现代应用数学丛书

信息论

〔日〕喜安善市 室贺三郎 著

上海科学技术出版社





統一书号 13119·480
定　　价 0.64 元

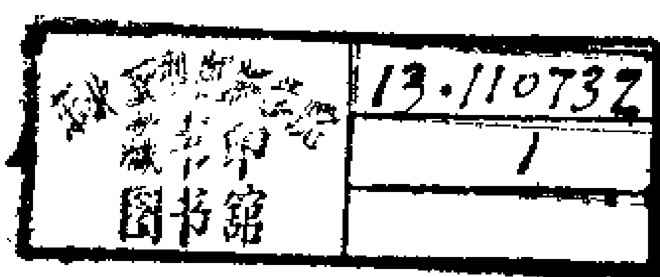
现代应用数学丛书

信　　息　　论

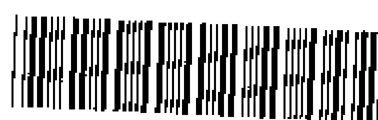
〔日〕喜安善市 室賀三郎 著

李　文　清　譯

陆　志　剛　校



上海科学技术出版社



內 容 提 要

本书是日本岩波书店出版的現代应用数学丛书之一的中譯本，介紹 Shannon 的信息傳遞理論。全书共七章，第一、二两章介紹信息的概念及信息量的数学表达式，第三章討論离散的信息源，四、五、六章分别討論无噪音及有噪音的离散信道及編碼方法，最后一章討論了連續的信息源。本书可供高等学校有关专业作为数学参考书，也可供工程师及研究人員参考。

現代应用数学丛书

信 息 論

原书名 情报理論

原著者 〔日〕喜安善市 室賀三郎

原出版者 岩波书店 1957

譯者 李文清

校者 鮎志剛

*

上海科学技术出版社出版

（上海瑞金二路450号）

上海市书刊出版业营业許可証出093号

新华书店上海发行所发行 各地新华书店經售

商务印书館上海厂印刷

*

开本 850×1168 1/32 印張 3 22/32 字数 85,000

1962年10月第1版 1962年10月第1次印刷

印数 1—4,500

統一书号：13119·480

定 价：（十四）0.64元

出版說明

这一套书是根据日本岩波书店出版的"現代应用数学讲座"翻譯而成。日文原书共15卷60册，分成A、B两組，各編有序号。現在把原来同一題目分成两册或三册的加以合并，整理成42种，不另分組編号，陆續翻譯出版。

这套书涉及的面很广，其內容都和現代科学技术密切有关，有一定参考价值。每一本书收集的資料都比較丰富，而叙述扼要，篇幅不多，有利于讀者以較短时間掌握有关学科的主要內容。虽然，这套书的某些观点不尽适合于我国的情况，但其方法可供参考。因此，翻譯出版这一套书，对我国学术界是有所助益的。

由于日文原书是1957年起以讲座形式陆續出版的，写作时間和篇幅的限制不可避免地会影响原作者对內容的处理，为了尽可能地减少这种影响，我們在每一譯本中，特請譯者或校閱者撰写序或后記，以介紹有关学科的最近发展状况，并对全书內容作一些評价，提出一些看法，結合我国情况补充一些資料文献，在文內过于簡略或不足的地方添加了必要的注釋和改正原书中存在的一些錯誤。希望这些工作能对讀者有所帮助。

承担翻譯和校閱的同志，为提高书籍的质量付出了巨大劳动，在此特致以誠摯的謝意。

欢迎讀者对本书提出批評和意見。

上海科学技术出版社

現代应用数学丛书

书名	原作者	譯者
代数学	弥永昌吉等	熊全淹
几何学*	矢野健太郎	孙澤瀛
复变函数	功力金二郎	刘书琴
集合·拓扑·測度*	河田敬义	賴英华
泛函分析*	吉田耕作	程其襄
广义函数*	岩村联	楊永芳
常微分方程	福原滿洲雄	張庆芳
偏微分方程*	南云道夫	錢端壮
特殊函数*	小谷正雄等	錢端壮
差分方程	福田武雄	穆鴻基
富里哀变换与拉普拉斯变换*	河田龙夫	錢端壮
变分法及其应用*	加藤敏夫	周怀生
李群論*	岩堀长庆	孙澤瀛
随机过程*	伊藤清	刘璋温
回轉群与对称群的应用	山内恭彦等	張質賢
結晶統計与代数	伏見康治	孙澤瀛
偏微分方程的应用	犬井鉄郎等	楊永芳
微分方程的近似解法	加藤敏夫等	王占瀛
数值計算法	森口繁一等	閻昌齡
量子力学中的数学方法*	胡永振一郎	周民强
工程力学系統*	近藤一夫等	刘亦珩
非綫性振动論	古屋茂	吕紹明
力学系与映射理論*	岩田义一	孙澤瀛
平面彈性論*	森口繁一	刘亦珩
有限变位彈性論 变形几何学*	山本善之 近藤一夫	刘亦珩
塑性論*	鷲津文一郎	刘亦珩
粘性流体理論*	谷一郎	刘亦珩
可压縮流体理論*	河村龙馬	刘亦珩
网絡理論	喜安善市等	陆志剛
自动控制論	喜安善市等	郑立林
回路拓扑学	近藤一夫	張鳴鏞
信息論	喜安善市等	李文清
推断統計理論	北川敏男	李賢平
統計分析*	森口繁一	刘璋温
实驗設計	增山元三郎	刘璋温
群体遺傳学的数学理論*	木村資生	刘祖洞
博奕論	官澤光一	張駿春
綫性規划	森口繁一	刘源張
經济理論中的数学方法	安井琢磨等	談祥柏
随机过程的应用*	河田龙夫	刘璋温
計算技术	高桥秀俊	姚晋
穿孔卡計算机	森口繁一	刘源張

注：有*者已在1962年7月以前出版。

目　录

第1章 序　論

本章首先敘述信息論的历史与发展过程，然后説明信息論中的信息、通信等基本概念。

§1 历史背景

人类和其他动物的明显区别是具有一种本能以外的通信手段。通信手段不限于声音，还有文字、繪画、雕刻、印刷等方面。其他动物不是沒有通信手段，但不象人类通信发展到那样的高度。准确、无誤和有效的通信是人类社会不可缺少的东西。虽然通信技术如光通信，电通信已經非常发达，但是信息通信的本质，以及妨害通信的噪音，在概率論的基础上加以理論的探討，还是到了廿世紀的中叶才开始的。

自然，現在完善的信息論，是人类多年关于通信本质研究的結果，是逐漸形成的理論，与通信技术密切相关，它的迅速发展則是由电通信的出現所促成的[1]。举一个熟知的例子，26 个英文字母中，E 是出現最頻繁的字母，Morse 电报就分配它短的碼，这样就縮短了英文信息傳送的平均时間(但是，妙的是用 Morse 碼对日文来說，就沒有这样的考慮)。其次，我們回顾一下自从电通信发展以来信息論的形成。

現在，家庭用的收音机使用着振幅調制❶，在 1922 年 Bell 电話研究所的 J. R. Carson 就指出，它是具有一定的頻带寬度，并且开始明确了边带的概念。其后，发明了現在长途电話使用单边

❶ 調制这一概念，参看文献[2]。对于不熟悉电气通信的讀者可以略去本节以下的內容，不会影响以后的理解。

带通信方式，討論了調頻，指出了調制指数小的調頻方式对寬頻带的带寬点的不实用性。此外，与信息論有关的研究是关于噪音波形的統計研究，这些都是最早的研究。

1924 年 H. Nyquist 及 K. Küpfmüller 两人独立地指出，电信信号的傳輸速率与信道带寬有比例关系。这种想法在 1928 年被 R. V. L. Hartley 推广了。Hartley 的想法也可以說是現在信息論的基础，他把信息考虑为代碼或是单語的序列，把它所代表的語义当做次要的而不予考虑。于是由 S 个代碼序列中选 N 个碼即构成 S^N 个可能的信息，他指出“信息量 H”定义为 $H=N\log S$ 是合理的。傳輸一定的信息量时，带寬及傳輸时間的乘积为常数，他論述了当带寬愈窄，傳輸时間愈长，带寬愈大，傳輸时間愈短。但 Hartley 的理論，沒有考虑到噪音和概率統計，現在的理論比他的想法可以說进步多了。

在 1940 年，岡田和藤木两人曾指出，信息的傳輸与能量的傳輸沒有直接关系，他們曾討論了 Hartley 信息量的各种具体例子。

信息論的进一步发展，是由于各种各样新通信方式的出現，作为通信的本质即信息的傳輸，要求在更广泛的基础上加以研究，又因为統計数学和妨碍通信的噪音理論的发展，信息論只有考虑到噪音及利用了数学統計的处理方法，才形成了理論。从社会环境来說，在第二次大战期間，为了制造新武器，动員了許多数学家、物理学家开展了以电子学为中心的軍事科学研究，这也是促使信息論发展的重要原因。

关于噪音的理論，V. D. Landon 在 1936 年及 1941 年曾討論过噪音的波形，指出它的峰値与有效値之比，不論噪音的带寬如何，常为 3.6 到 4. K. Fränz 在 1940 年討論了射頻接收机的檢波器問題，1943 年 P. A. Mann 及 1947 年瀧保夫对噪音波形作了細致討論，都是关于噪音波形的理論。1945 年及 1948 年 Bell

研究所的 S. O. Rice 曾討論了随机噪音的統計性质，并对噪音与正弦波迭加的情况，作了詳細的数学分析[3]。

关于通信方式的发展，E. H. Armstrong 在 1936 年最初使用了調頻通信装置，調制指数很大时，带寬虽然变大，但有抑制噪音的作用。随着在第二次大战时期，脉冲技术的急速发展，脉冲幅度調制(PAM)，脉冲寬度調制(PWM)，脉冲頻率調制(PFM)，脉冲位置調制(PPM)等各种通信方式发展起来，其中特别重要的是对噪音抑制效果最大的脉冲編碼調制(PCM)。早在 1939 年，H. Reeves 曾对他的原理申請专利，1948 年 Bell 电話研究所克服了技术上的困难完成了制造。此种脉冲編碼調制，后来对信息論的形成充当了重要角色。

在二次大战时期和战后，电子計算机和各种武器的自动控制的蓬勃发展，刺激了英美諸国在这方面的研究，在信息傳輸形式的广泛基础上，把噪音也考虑在內的研究才探明了信息通信的本质。

1946 年 D. Gabor 及 1940 年今堀，討論了信号的长度、頻率的不确定性。其后 Szilard, Weiner, Brillouin 等人討論了信息量与热力学第二定律的关系，高桥又討論了信息論与統計力学的关系。

Wiener 在战时曾研究抑制噪音的滤波器設計問題[4]。把信号及噪音看做随机过程，做出了信号波形与信号上重迭噪音波形之間最小均方誤差的滤波器設計。所化的代价是时間的推迟，推迟愈多，錯誤概率愈小。他还研究了預測器，給出了比輸入信号早一段时間的信号值具有均方誤差的預測器設計。时間愈提前，錯誤概率愈大。

此后，李郁荣用电子綫路实現了 Wiener 相关器和預測器，能把信号噪音比为 −20 分貝的埋藏在噪音中的周期信号用相关器提取出来，能够改善信号噪音比达 +20 分貝左右[5]。

第二次世界大战开始时，Wiener 与当时哈佛大学的医学家 A. Rosenblueth 試圖把現代各学科中通信及自动控制的基本問題綜合成一門新的学問，他們在很长一段时期內，同各方面的学者举行了学术討論会，終于在 1947 年命名为控制論 (Cybernetics)[6]。此种气氛和刺激对于近代信息論的发展有不可否认的影响。在这种气氛中，除 Wiener 的相关器及預測器外，还对信息量提出了新的定义，这就是 R. A. Fisher 的不同定义[7]，用了統計力学的熵的形式。

1948 年 Bell 电話研究所的 C. E. Shannon 提出了最完善的統一的信息論[8]。即把 Hartley 的信息量在概率論更广泛的基础上加以定义，同时发展了有噪音时信息傳輸的理論。比他較早一些时間，W. G. Tuller 曾討論了有噪音干扰的信道的信息傳輸量問題[9]。Shannon 利用熵的形式，导入了通信容量这一新的重要概念。由于 Shannon 划时代的貢献，使信息論很快地形成了雛形。其后，經过許多学者的努力，才使这个理論漸趋丰富与完善。

通常"通信理論"与"信息論"几乎是同义語。重要的区別在于后者与控制論有相同的含义，而前者是控制論中与通信有关的一部分而已。从另一种观点說，所謂"信息論"，即是通信系統中用近代数理統計学的方法探究信息傳輸問題的科学。关于信息本质的科学从广义来說，除調制問題、傳輸理論、頻譜分析問題外，有人从物理学的角度认为热力学、量子力学也有信息論的本质問題。这本小册子以 Shannon 的研究为中心說明信息理論，可以认为信息本质是最基本的，由这种角度写的信息論专书在国内外还不多。

§2 信息是什么？

"信息" (information)，在辞典(如日文广辞苑)上的意义是所观察事物的知識，"通信"是两人以上互相交换信息或意志。如果

不是交换，就叫做单方傳輸了。但在定义信息量时，“信息”、“通信”的意义就与通常的涵意不同，必须注意使用了抽象化的概念。“通信”放在下一节討論，本节只叙述“信息”。

同样一件特定的通信内容，对不同的接收者来說，常引起不同的情感。例如1956年埃及納賽尔总統宣布苏彝士运河国有化时，英法两国感到挫伤，阿拉伯諸国莫不称快。从信息論来說，对不同的人，由于接收的意义不同，不能看做与信息量有关。信息的接受是把事情本身、时間和内容都撇除，使之抽象化，在各种情况下，問题在于代表不同的知識。从这个角度看来，在苏彝士运河国有化問题上，重要的是事情发生以后苏联对中东的影响，英国在地中海霸权的丧失，納賽尔总統的盛衰等事情发生的可能性。假定可以猜到納賽尔的盛衰，此外沒有其他的可能性，那末苏彝士运河的国有化就算不得什么大信息。若一切的可能性都相等，那就无从猜測国际形势的变化，于是这就成为国际形势中的一个大信息。

收音机中人們喜欢收听的“20个屏”[1]恰好适合信息量的概念。猜謎的人提出20个問题获得信息，来判断謎底是什么。后提的問题应和以前提的尽量不同，并且提的問题要最大可能地与謎底有关，这样才能有效地利用这20个信息所构成的信息量。

不只是“簡单地有多少个可能性”，而更重要的是“这样或那样事件有多少程度的可能性”。Hartley 只是由“有多少个可能性”这一点出发建立信息論，而 Shannon 理論的出发点是“这样或那样的事件有多少程度的可能性”，这里使用了統計概率論，这是近代信息論发展的基础。

信息論最重要的一点是，当通信文已經知道的时候，注重的不在此特定的通信文本身的意义及内容，而是此特定通信文以外可

---

[1] “20个屏”是日本民間流行的一种猜謎形式。猜的人可以向出“謎”的人提出廿个問題，故称20个屏。出謎的人要誠实回答問题。——譯者注

能产生多少种通信文，又各通信文实际上有多少出現概率。換言之，重要的不是“我們述說的事物”，而是“我們所能述說事物的大概程度”。

我們用語言交换信息，即进行通信，不論使用自然語或人工語，如果文章的长度相同，則含有不同意义的单字越多的文章，它的信息量愈大。例如用48个日文字母写的文章，与夹杂着难解的汉字所写的文章相比較，不論讀和写，前者因为字母少因之很容易。反之，后者的信息量較大。然而夹杂着汉字的文章，汉字数以万計，但大部分的汉字出現的可能性不大，如果限制使用一部分汉字，則不費很大的劳力也能写或讀。

在电报中所用的代碼是一种广义的人工語，亦有相同的情况，单位碼的种类多的通信系統，同一时間所送的信息比較多。例如对于脉冲編碼調制，不管其技术上的难易，用三元碼通信比用二元碼通信，在一定时間內，会送出更多的信息。

§3 如何进行通信

我們如何进行信息通信呢？ 例如我們給远方的人打电报，把电文写在紙上交給邮电局，电报員把文字变成 Morse 电碼电流(或电傳打字电碼电流)，通过綫路傳送出去。在綫路中途有时受到噪音的干扰而产生錯字，收信局再把电碼电流譯成文字，印刷之后，送交收信人。

又如长途电話，打电話人的声音經过炭质送話器变成声音电流送到长途电話局，換成載波电流經过很长的綫路，到达目的地的电話局，再变換成声音电流，經过听电話人的电話耳机把声音电流变成声音，再傳到听电話的人。此种情况下，在綫路、交換机、中继器內可能混入噪音。

信息論是把以上所述各个实际出現的情况加以抽象化、理想

化，成为通信系統模型的数学理論。由图 3.1 看出，这个模型分成下列五部分。

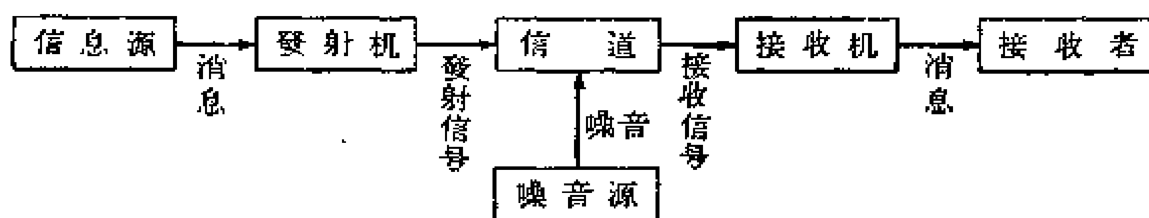


图 3.1

(1) **信息源** (information source) 产生要傳輸信息的来源称为信息源，从信息源产生的信息叫做消息 (message)。如打电报的人的大脑卽是信息源，写在紙上的文字卽是消息。在电通信中，电报、电傳打字中所用的字母序列(包括标点在內)、电話或无綫电广播的声压变化的时間函数、电視中对应的时間及空間坐标的函数、黑白电視中的二元坐标系所表示的点 (x, y) 处光强度的时間 t 的函数 $f(x, y, t)$，就都是消息的例子。

(2) **发射机** (transmitter) 把消息变成适合于信道傳輸的信号的装置叫发射机。例如 Morse 电报的情况，发射机就是把电文变成点、划、空白等电碼序列所用的电报机，在电話的情况下，发射机是把声压按比例变成电流的炭质送話器。但是，从信息論看来，声压、电流等表示信息的物理要素間的变化，不是重要的事，重要的是在通报时能把字母或电碼序列变成别的电碼序列，从这种性质着眼，发射机有时亦称为变换器。例如在 Morse 电报情况下，字母变换成电流的技术問題，从信息論的立場看来并不重要，重要的是字母序列的适当限制，例如不使字母空白連續发生两次，不使单語空白与字母空白連續发生的限制，卽具有一定时間长度的电碼序列的变換是很重要的。

(3) **信道** (channel) 是信号由发射机傳輸到接收机时所使用的媒介。例如电綫、同軸电纜、一定频带寬度的无綫电波，在光

通信中的光綫等都是媒介。信号在信道傳送的中途，能受到噪音的干扰，由发射机发出的信号，不一定正确地傳送到接收机。此外，信号一般受到的失真，在接收机中普通是能够还原的，可是，本质上妨碍信息傳送的噪音却与这种失真是不同的。

(4) **接收机** (receiver) 是把信道傳送来的信号与噪音的混合接收下来，把原来的消息加以还原，它进行和发射机相反的操作。此时电碼的变换器的作用非常重要。

(5) **接收者**(destination) 即是接收消息的人或物。

以上的通信的模型，不仅适用于电气通信系統，即使象神經系統等等广义的通信系統也是适用的，它們只有形式上的差別，根本上是相同的。信息交換是由相对方向的两个通信系統組成，各个方向的通信系統即是上述的模型。这模型对观察自然界也是形成一个信息观察系統，并且在某种程度上也适用于社会现象中各种形式的信息傳輸。

为了数学上处理的方便，数量通常可分为离散的及連續的两种情况。电报的电文是由离散的信息源发生的离散的消息，Morse电报的点、划信号是离散的信号，其傳送的信道亦称为离散的。但是电話的消息和信号都是連續的，它的信息源、信道皆称为連續的。但处理的数量，有时同时有离散的及連續的，即所謂混合型的，这在数学上很困难，实际上也不重要。

第2章 信息量与熵

关于信息量的数学表达式，根据信息量所应满足的条件，我們自然地导出了熵的形式，然后再闡述熵的性质。

§4 信息量的数学表达式

前章闡述了信息的本质，但信息量的数学公式如何表示呢？以后将要談到，由信息源产生的消息一般可以用 Markoff 过程表示，此处只談构造最簡单的信息源，即消息完全不受过去影响的信息源。

为了理解上的便利，現在用概率論的語言加以說明。設有 n 个不同事件（即信息論中所謂消息使用的字母），从中任选一个，选擇反复进行构成偶然事体（即信息源），且前一选擇与后一选擇无关，各事件出現的概率为 $p_1, p_2, \cdots, p_n$. 此种事体的信息量以 $H(p_1, p_2, \cdots, p_n)$ 表之，希望有下列性质：

（i） 信息量 $H(p_1, p_2, \cdots, p_n)$ 为 $p_1, p_2, \cdots, p_n$ 各自变量的連續函数。

（ii） 当 $p_i\,(i=1, 2, \cdots, n)$ 都是 $\frac{1}{n}$ 时，H 是 n 的单調递增函数。

(iii) 一个事件的选擇分成两个步驟时，未分之前的 H 即是既分之后 H 的加权和。

这个(iii)的意义如图 4.1(*a*) 所示，三个事件的概率为 $p_1=\frac{1}{2}$, $p_2=\frac{1}{3}$, $p_3=\frac{1}{6}$. 变成图 4.1(*b*) 时，先取概率为 $\frac{1}{2}$ 的两个事件中的任一个，第二阶段分別选擇概率为 $\frac{2}{3}$, $\frac{1}{3}$ 的事件。选擇三个事

件的概率与分两个步驟无关，各个概率最后是同样的，此时希望下式成立：

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)=H\left(\frac{1}{2}, \frac{1}{2}\right)+\frac{1}{2}H\left(\frac{2}{3}, \frac{1}{3}\right).$$

加权系数 $\frac{1}{2}$ 是由于在第二选擇阶段，从时間上說，只成一半了。

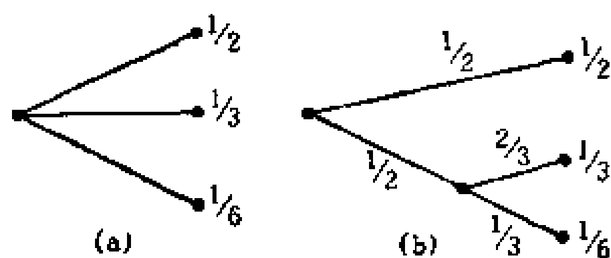


图 4.1

满足上列三个条件的H，必然是統計力学的熵的形式。首先考虑 p_i $(i=1, 2, \cdots, n)$ 都是相等的情况，置

$$H\left(\frac{1}{n}, \cdots, \frac{1}{n}\right)=A(n).$$

由条件(iii)，等概率下发生的 s^α 个事件，等于概率为 $\frac{1}{s}$ 的 s 个事件分解为 α 阶段重复的选擇，如图 4.2(a)，于是得出

$$A(s^\alpha)=\alpha A(s).$$

同样对于与 s 不同的 t，得

$$A(t^\beta)=\beta A(t).$$

当 β 适当大时，选取满足下式的 α：

$$s^\alpha \leqslant t^\beta < s^{\alpha+1}.$$

取对数，以 $\beta \log s$ 除它，得

$$\frac{\alpha}{\beta} \leqslant \frac{\log t}{\log s} < \frac{\alpha}{\beta}+\frac{1}{\beta}. \tag{4.1}$$

另一方面，由条件(ii)，得公式

$$\alpha A(s) \leqslant \beta A(t) < (\alpha+1) A(s),$$

上式除以 $\beta A(s)$ 得

$$\frac{\alpha}{\beta} \leqslant \frac{A(t)}{A(s)} < \frac{\alpha}{\beta}+\frac{1}{\beta}.$$

把上式和(4.1)式合起来，設 ε 为任意小正数，α, s, t 相当大时，取

充分大的 β，可得

$$\left|\frac{A(t)}{A(s)}-\frac{\log t}{\log s}\right|<\varepsilon. \tag{4.2}$$

换言之，对于充分大的 β，我們有

$$A(t)=-K\log t. \tag{4.3}$$

由条件(ii)知 K 为正数。

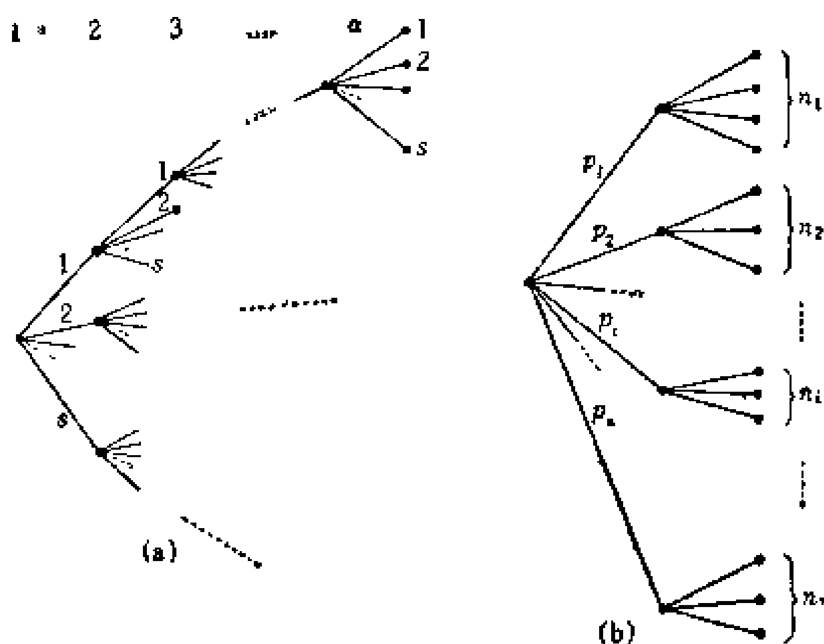


图 4.2

令設 n 个事件的概率 $p_1, p_2, \cdots, p_n$ 都是有理数，n_i 为整数，則可取

$$p_i=\frac{n_i}{\sum_{i=1}^{n} n_i}.$$

即 $\sum_{i=1}^{n} n_i$ 个事件以相等的概率选擇时，如图 4.2(b)，n 个事件的选擇概率为 $p_1, p_2, \cdots, p_n$，在第 2 阶段就可以等概率分解各 n_i 个事件。所以由条件(iii)和前半的証明，这个选擇的信息量为

$$K\log\sum_{i=1}^{n} n_i=H(p_1, \cdots, p_n)+K\sum_{i=1}^{n} p_i\log n_i.$$

于是得

$$H(p_1, \cdots, p_n) = K\left[\sum_{i=1}^{n} p_i \log \sum_{i=1}^{n} n_i - \sum_{i=1}^{n} p_i \log n_i\right]$$

$$= -K\sum_{i=1}^{n} p_i \log \frac{n_i}{\sum_{i=1}^{n} n_i} = -K\sum_{i=1}^{n} p_i \log p_i. \quad (4.4)$$

由条件(i)知H是p_i的連續函数，有理数稠密地存在在区間内，p_i为无理数时可以用有理数近似，所以由条件(i)，对无理数的p_i，(4.4)式同样成立。常数K是任意的，当单位的值一經选定就規定好了。因此下列定理成立。

定理4.1 滿足(i)，(ii)，(iii)条件的信息量可用下式且仅用下式表示：

$$H(p_1, \cdots, p_n) = -K\sum_{i=1}^{n} p_i \log p_i, \quad (4.5)$$

其中K为正的常数。

这个H与統計力学中熵的形式完全一样。它表示信息源(偶然事体)中一个字母(事件)平均負担的信息量，同时也是信息源中特定消息选擇的自由度，或是表示統計推斷的不确定性。熵是信息論中最重要的表示式。定理中，三个条件对以下的理論不是必要的，但信息量表示式的合理性十分重要。

与統計力学一样，上列定理中的H叫概率$p_1, p_2, \cdots, p_n$的**熵**。以下用x表示随机变数，而以$H(x)$表示其熵。

当$K=1$并以2为对数底时，在(4.5)中信息量的单位叫**必特**(bit)❶。这对于以下所述的继电器、触发器等电子計算机所用的电子綫路的穩定动作状态的两个位置而言，也是适合的。但是在解析計算上取e为底比較方便，此时称其单位为**納特**(nat)。在数值計算中，采用常用对数，这时单位称为**十进位单位**。今后若不需

❶ J. W. Tukey 建議，为了使数学上的二进位数与信息量的单位必特相区别，单位用必尼特(binit)[10]。

特别指定对数的底，即取任意底都可以时，就简单用必特表示单位熵，或者不标明单位。

在实际应用上，对数以 2 为底非常便利。当 $p_i=2^{-m}(i=1,2,\cdots,n$, 且 $n=2^m)$ 时，其熵就是

$$H=-\sum 2^{-m}\log_2 2^{-m}=m\ (\text{必特})。\tag{4.6}$$

取对数底为 2 时，在"是"与"否"之間选一个相当于在最簡单的信息基础上的信息量，二者之一选一个，反复进行 m 段，便能在 2^m 个中指定出一个。"是"与"否"又可对应于 m 段继电器的上下接点，根据继电器的連接，可以指定一个状态(图 4.3 表示 $n=2^2=4$ 的例)。設想这里各段的继电器負担着一个信息，全体共 m 倍。

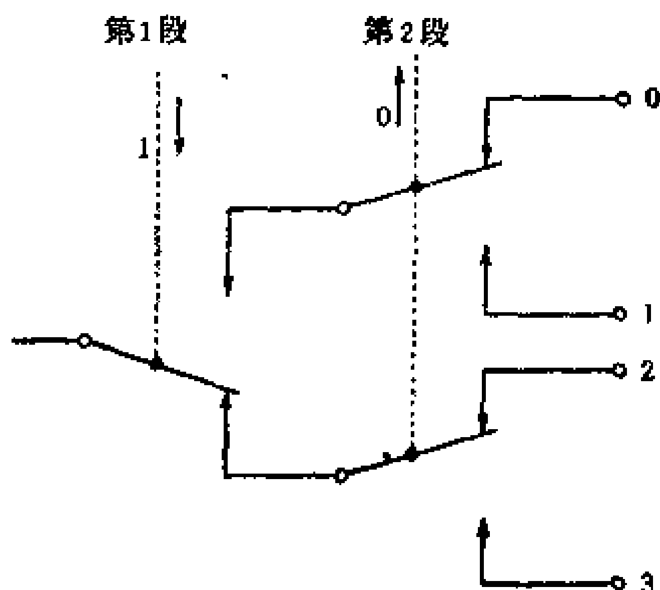


图 4.3

各接点的上位置表示 0，下位置表示 1，m 段的接点能表示二进位 m 个数碼所表示的数，并能在 0 到 2^m-1 共 2^m 个端子中唯一指定某一端子(图 4.3 中 2 段接点的位置表示二进位数的 10，它在十进位数中即是 2，图中标碼为"2"的端子即所指定的端子)。因为它們都是相同的，故由 2^m 个中选取一个的信息量叫 m 必特的信息。

在信息論中，关于单位的选擇很重要，即在单位时間內，偶然事体选擇的次数問題。現在假定每次选擇的継續时間不同，設在单位时間內平均发生的次数为 γ，用这 γ 乘(4.5)式❶(令 $K=1$，其单位取必特)，得一新的单位"必特/秒"，写成

$$H'=\gamma H.\ (\text{必特/秒})❶\tag{4.7}$$

❶ 此处原书为"用这 γ 除(4.5)式"; "$H'=H/\gamma$".——校者注

这个量称为偶然事体反复选擇时在单位时間内的信息量，它具有动态的概念，与(4.5)不含时間的靜态概念不同，(4.5)式的单位簡单地只写"必特"，以后所述的事体，除了信息源的字母选擇以外，又考虑到信道上构成信号的代碼，不管是字母或是代碼，为了清楚起見，常写成"必特/字母"或"必特/代碼"❶。但它們沒有时間概念，必須与(4.7)的单位有严格的区别。

§5 离散随机变数的熵的性质

假定离散随机变数經过一段时間后取到 n 个不同的互相独立的值，則其熵如§4所述为

$$H=-\sum_{i=1}^{n}p_i\log_2 p_i \quad (\text{必特})。\tag{5.1}$$

本节将对多个随机变数推广熵的定义，并說明熵的一些性质❷。

(i) 变数为两个时，

$$H=-\{p\log_2 p+(1-p)\log_2(1-p)\} \quad (\text{必特})，\tag{5.2}$$

这如图5.1所示。

(ii) 当 p_i 仅有一个，且取值为1，其他 p_i 都是零时，(5.1)的熵 H 才取值为0，即只有一个事体被确定时 H 等于0，除此以外 H 恒为正值。

(iii) 当所有的 p_i 取值 $\frac{1}{n}$ 时，H 达到最大值 $\log_2 n$。物理意义为相当于最不确定的情况。

(iv) 設有两个离散的随机变数 x 及 y，經过一定的时間，x 取 n 个相异的值，y 取 m 个相异的值，彼此不一定独立。两个变数 x 及 y 取 i 及 j 的联合分布概率为 $p(i,j)$，这 x 及 y 的联合分布

---

❶ 在以后，偶然事体的事件考虑为代碼序列时，記其单位为"必特/代碼序列"，一个代碼序列包含有 k 个代碼时，以 k 除之，可得"必特/代碼"。

❷ 以下的証明不困难，讀者可以自己試作一下。



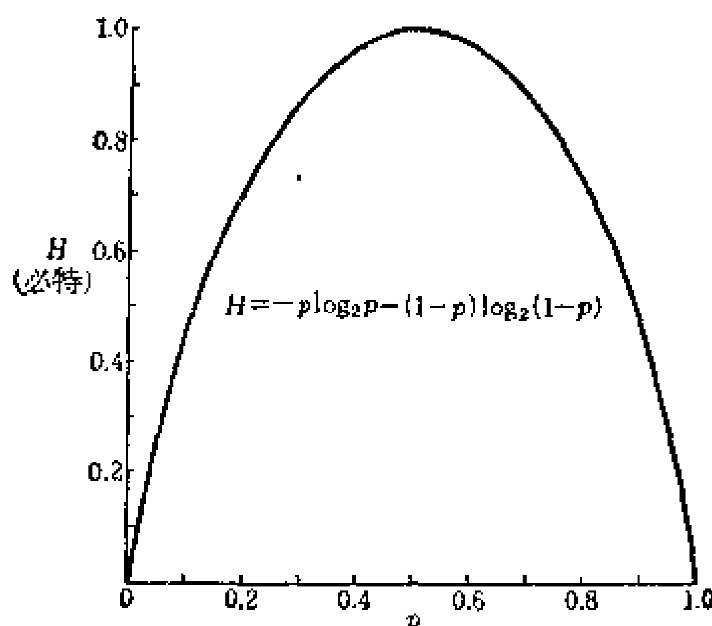


图 5.1

的熵，或联合熵定义为

$$H(x, y)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(i, j)\log p(i, j). \tag{5.3}$$

另一方面，x 及 y 各个单独的熵不外是 (5.1) 的形式，用联合分布概率 $p(i, j)$ 改写时，得出

$$\left.\begin{aligned}H(x)&=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(i, j)\log\sum_{j=1}^{m}p(i, j),\\H(y)&=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(i, j)\log\sum_{i=1}^{n}p(i, j).\end{aligned}\right\} \tag{5.4}$$

容易証明，联合熵满足下列不等式：

$$H(x, y)\leqslant H(x)+H(y). \tag{5.5}$$

当 x 与 y 彼此独立时，即 $p(i, j)=p(i)\cdot p(j)$ 时，等号成立。

(v) 概率 $p_1, p_2, \cdots, p_n$ 向着彼此相等的趋势变化时，熵 H 逐渐变大。例如，$p_1<p_2$ 时，若 p_1 增大，p_2 减少，則 $|p_1-p_2|$ 比以前小，于是 H 增大。一般地說，对概率 $p_1, p_2, \cdots, p_n$ 施以平均化的变换时，

$$p_i'=\sum_{j}a_{ij}p_j \quad (i=1, 2, \cdots, n),$$

且
$$\sum_i a_{ij}=\sum_j a_{ij}=1 \text{ 及 } a_{ij}\geqslant 0.$$
在这种变换下，除了簡单的置换以外（此时 H 不增大也不减少），熵 H 会增大。

这与統計力学的 H 定理所述的熵的增大的法則是相同的，証明也是利用熵的凸函数性质，方法相同。

(vi) 以下定义两个随机变数的熵。設 x 及 y 为两个随机变数，当 x 取值 i 时，y 取值 j 的条件概率为 $p_i(j)$. 此时，取 x 的各个値对应的 y 的熵，即 x 各値的概率的加权平均，可得出对应于 x 的 y 的条件熵。即
$$H_x(y)=-\sum_{i=1}^{n}P_i\sum_{j=1}^{m}p_i(j)\log p_i(j), \tag{5.6}$$
其中 P_i 为 x 取値 i 的概率。此即已知 x 时 y 的平均不确定性。由概率論得关系式如下：
$$\left.\begin{aligned} p_i(j)&=\frac{p(i,j)}{P_i},\\ P_i&=\sum_{j=1}^{m}p(i,j).\end{aligned}\right\}$$
利用上式，(5.6)式可改写为
$$\begin{aligned}H_x(y)&=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(i,j)\log p(i,j)+\sum_{i=1}^{n}\sum_{j=1}^{m}p(i,j)\log\sum_{j=1}^{m}p(i,j)\\&=H(x,y)-H(x),\end{aligned}$$
或写成
$$H(x,y)=H(x)+H_x(y). \tag{5.7}$$

(vii) 由(5.5)及(5.7)式得出
$$H(x)+H(y)\geqslant H(x,y)=H(x)+H_x(y),$$
即
$$H(y)\geqslant H_x(y). \tag{5.8}$$
此式説明当 x 为已知时，y 的不确定性不会增大。設 x 与 y 是相互独立的，则上式等号成立。

§6 連續随机变数的熵的性质

声音經过微音器后的电压或电視信号的电压这类有連續值的連續随机变数的熵，可以和§5离散的变数同样地定义出来。

随机变数 x 的概率密度为 $p(x)$ 时，其熵的定义即是 Σ 号换为积分号的(4.5)式：

$$H=-\int_{-\infty}^{\infty}p(x)\log p(x)\,dx. \tag{6.1}$$

当对数底为2时，H 的单位为必特。

設 n 个随机变数的联合分布概率密度为 $p(x_1, x_2, \cdots, x_n)$，其**联合分布熵**即**联合熵**定义为

$$H=-\int\cdots\int p(x_1,x_2,\cdots,x_n)\log p(x_1,x_2,\cdots,x_n)\,dx_1\cdots dx_n. \tag{6.2}$$

在两个变数 x 及 y 的特殊情况下，其联合熵和**条件熵**各为

$$H(x, y)=-\iint p(x, y)\log p(x, y)\,dx\,dy, \tag{6.3}$$

$$H_x(y)=-\iint p(x, y)\log\frac{p(x,y)}{p(x)}\,dx\,dy, \tag{6.4}$$

$$H_y(x)=-\iint p(x, y)\log\frac{p(x,y)}{p(y)}\,dx\,dy, \tag{6.5}$$

其中

$$p(x)=\int p(x, y)\,dy, \quad p(y)=\int p(x, y)\,dx. \tag{6.6}$$

以上五个式子中的 x 及 y 对应的概率密度在(6.2)式中同样可看做 n 个变数的联合分布概率密度。即 x 及 y 各表示变数組 $x_1, x_2, \cdots, x_n$; $y_1, y_2, \cdots, y_n$ 时，(6.3)至(6.6)式都能从一元情况推广到多元情况。此时下述的熵的性质与§5的情况大部分是相同的*。

* 証明不困难，讀者可以自己练习一下。

（i）設 x 在 Euclid 空間中体积为 V 的子空間内变动，当 $p(x)$ 在此空間内等于常数 $\frac{1}{V}$ 时，$H(x)$ 达到最大値 $\log V$。

（ii）
$$H(x, y) \leqslant H(x) + H(y), \tag{6.7}$$
但当 x 与 y 相互独立时，等号成立。

（iii）概率密度 $p(x)$ 施以加权平均时，得

$$p'(y) = \int a(x, y) p(x) dx,$$

但
$$\int a(x, y) dx = \int a(x, y) dy = 1, \quad a(x, y) \geqslant 0.$$

由概率密度 $p'(y)$ 所得的熵不会比原来的 $p(x)$ 所得的熵为少。

（iv）
$$H(x, y) = H(x) + H_x(y) = H(y) + H_y(x)$$

及
$$H_x(y) \leqslant H(y). \tag{6.8}$$

（v）当联合分布概率密度 $p(x_1, x_2, \cdots, x_n)$ 的二阶矩

$$A_{ij} = \int \cdots \int x_i x_j p(x_1, \cdots, x_n) dx_1 \cdots dx_n \tag{6.9}$$

固定之后，达到最大熵的是具有 n 維正态分布的二阶矩 A_{ij}，其概率密度为

$$p(x_1, x_2, \cdots, x_n) = \sqrt{\frac{|a_{ij}|}{(2\pi)^n}} \exp\left(-\frac{1}{2} \sum_{i,j} a_{ij} x_i x_j\right), \tag{6.10}$$

其中 A_{ij} 是矩陣 (a_{ij}) 的逆矩陣的元，$|a_{ij}|$ 表示矩陣 (a_{ij}) 的行列式。这个最大熵等于

$$H = \log (2\pi e)^{\frac{n}{2}} |a_{ij}|^{-\frac{1}{2}}. \tag{6.11}$$

特别，当 $n=1$ 时情况比較簡单，兹加以詳細論証。設 x 的标准偏差为一常数 σ，則最大熵的概率分布密度 $p(x)$ 为正态分布。要証明这一点，即求在下列約束条件下，

$$\left.\begin{aligned} \sigma^2 &= \int x^2 p(x) dx, \\ \int p(x) dx &= 1, \end{aligned}\right\} \tag{6.12}$$

使以下的熵:

$$H(x)=-\int p(x)\log p(x)\,dx \tag{6.13}$$

达到最大。根据变分法,亦就是使

$$\int[-p(x)\log p(x)+\lambda x^2 p(x)+\mu p(x)]\,dx \tag{6.14}$$

达到最大,式中 λ, μ 为待定乘数。为了上式达到最大,要求

$$-1-\log p(x)+\lambda x^2+\mu=0, \tag{6.15}$$

再由約束条件决定 λ 和 μ, 結果得出

$$p(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\exp(-x^2/2\sigma^2). \tag{6.16}$$

此时熵为

$$H(x)=\log\sqrt{2\pi e}\cdot\sigma. \tag{6.17}$$

(vi) 設 $x\leqslant 0$ 时, $p(x)=0$, 且設

$$a=\int_0^\infty xp(x)\,dx \tag{6.18}$$

为一約束条件,則最大熵的概率密度为

$$p(x)=\frac{1}{a}\exp(-x/a), \tag{6.19}$$

这个最大熵为

$$H(x)=\log e\,a. \tag{6.20}$$

(vii) 連續随机变数与离散随机变数的熵具有重要的区别。

在离散的情况下,当概率确定时,熵就唯一决定了。对于連續的情况,即使概率密度已知时,熵不能唯一决定,因为坐标系不同。設坐标系 $x_1, \cdots, x_n$ 对应的熵为 $H(x)$, 换为新坐标 $y_1,\cdots,y_n$ 时,其熵为

$$H(y)=H(x)-\int\cdots\int p(x_1,\cdots,x_n)\log J\left(\frac{x}{y}\right)dx_1\cdots dx_n. \tag{6.21}$$

其中 $J\left(\frac{x}{y}\right)$ 为坐标变换的 Jacobi 行列式。换言之,給出了对給

定的坐标系微小容量 $dx_1\cdots dx_n$ 均等加权平均所表示的无规则性的度量。

因此，連續随机变数的熵因坐标系不同而不同，这有不方便之处。但以后将証明，“傳輸速率及通信容量两个重要概念，可以定义成两个熵的差，这两个熵的差具有不因坐标系而变的性质”。所以熵的概念仍有重要意义。上式的熵，因坐标的选擇，有时为負，但傳輸速率及通信容量是非負的。

(viii) 考虑坐标变换是綫性变换

$$y_i=\sum_{i=1}^{n}a_{ij}x_i \tag{6.22}$$

的特殊情况。这时，Jacobi 行列式为 $|a_{ij}|$ 的倒数，于是

$$H(y)=H(x)+\log|a_{ij}|. \tag{6.23}$$

在坐标系旋轉的情况下(或保測变换)，$J=1$，則 $H(y)=H(x)$。

第3章 离散的信息源

本章先闡述在信息源及信道的数学表示上起本质作用的Markoff过程，特别是遍历过程的数学預备知識。然后叙述作为随机过程的信息源的特点，表示信息源所具有的信息量及信息的冗长度可以使用熵加以測度。

§7 Markoff 过程及 Shannon 線图[11][12]

在信息論的叙述中，随机过程論的知識是不可缺少的，作为准备，玆概述如下。

先設 x_t 表示一个受概率法則支配的事物在 t 时刻实现的事件。这样以时間为参数的随机变数集合 $\{x_t\}$ 即：$\cdots, x_0, x_1, x_2, \cdots$ 所成的序列叫做随机过程 (Stochastic process)。在古典概率論中，常把 t 考虑为有限个自然数值，我們这里 t 是无限个自然数或取連續值。随机过程的样本即实現(或事件)的序列，特别称为**时間序列** (time series)。

由信息源发生的消息中的字母或信道中傳輸的电信号的波形具有非离散的連續振幅，它們也可看做随机过程。即 x_t 不仅是由信息源发生的消息在时間 t 用字母所实現出来的事件，也可以是电压振幅值的数量。

随机过程 $\{x_t\}$ 的現在事件 x_t，如果只与从現在向后退 n 个时刻的事件有关，而与再以上的事件无关，則称这随机过程为 **n 重 Markoff 过程**。特别簡单的1重 Markoff 过程，叫做**单純 Markoff 过程**。

在 n 重 Markoff 过程情况下，可用 n 維矢量 $\hat{x}_t = (x_t, x_{t+1}, \cdots,$

x_{t+n-1}) 代替一个随机变数 x_t，而对于新的矢量随机过程 $\{\hat{x}_t\}$，只要知道前一时刻的事件 $\hat{x}_{t-1}$ 就够了。换言之，矢量 $\hat{x}_{t-1}$ 的第一个坐标是 x_{t-1}，其他的坐标都与 $\hat{x}_t$ 内的坐标相同，所以相继的 $n+1$ 个事件都可知道。这样，使用矢量随机变数就可把 n 重 Markoff 过程归結为单純 Markoff 过程。因此，Markoff 过程一般是指单純的 Markoff 过程而言。

当随机变数只取有限个离散值时，n 重 Markoff 过程可以用綫图表示，这样的綫图用来表示信息源或信道是很方便的。

現在用綫图考虑有 n 重 Markoff 过程性质的信息源[13]。这过程的各个随机变数 x_t 能取 N 个字母 L_1, L_2, …, L_N 中的一个字母时，則矢量(x_t, x_{t+1}, …, x_{t+n-1})至多有 N^n 种組合，这些矢量都称为**信息源的状态**。这不是随机过程 $\{x_t\}$，而与矢量随机过程 $\{\hat{x}_t\}$ 所取的事件相当。矢量具有什么样的坐标組合是由信息源的固有性质决定的。这些状态之間，若用方向綫連結，信息源由一个状态迁移到另一个状态时，其連結綫所对应的預定的字母可考虑作信息源的发生处。卽由任一状态 (L_{i_0}, L_{i_1}, …, $L_{i_{n-1}}$) 迁移到其他状态至多有 n 个可能性，卽(L_{i_1}, L_{i_2}, …, $L_{i_{n-1}}$, L_k)，其中 $k=1$, 2, …, N. 于是这个状态变动一次，其相应的信息源便出現各个字母L_k。例如图 7.1 的 (a)，(b)及(c) 各表示两个字母的单純 Markoff 过程、2 重及 3 重 Markoff 过程的綫图；图(d) 表示三个字母的单純 Markoff 过程的綫图，图 (e) 則是四个字母的单純 Markoff 过程的綫图。

单是綫图沒有多大意义，在此种表示信息源的綫图中，由任一状态 E_i 沿着方向綫迁移到另一状态 E_j，考虑迁移概率时，才有重要意义。这綫图中，一切方向綫有关的迁移概率与信息源任意开始状态的初期概率分布完全規定了 Markoff 过程。特別，方向綫的迁移概率为零时很重要。就象图 7.1 的綫图，形式上虽然画着

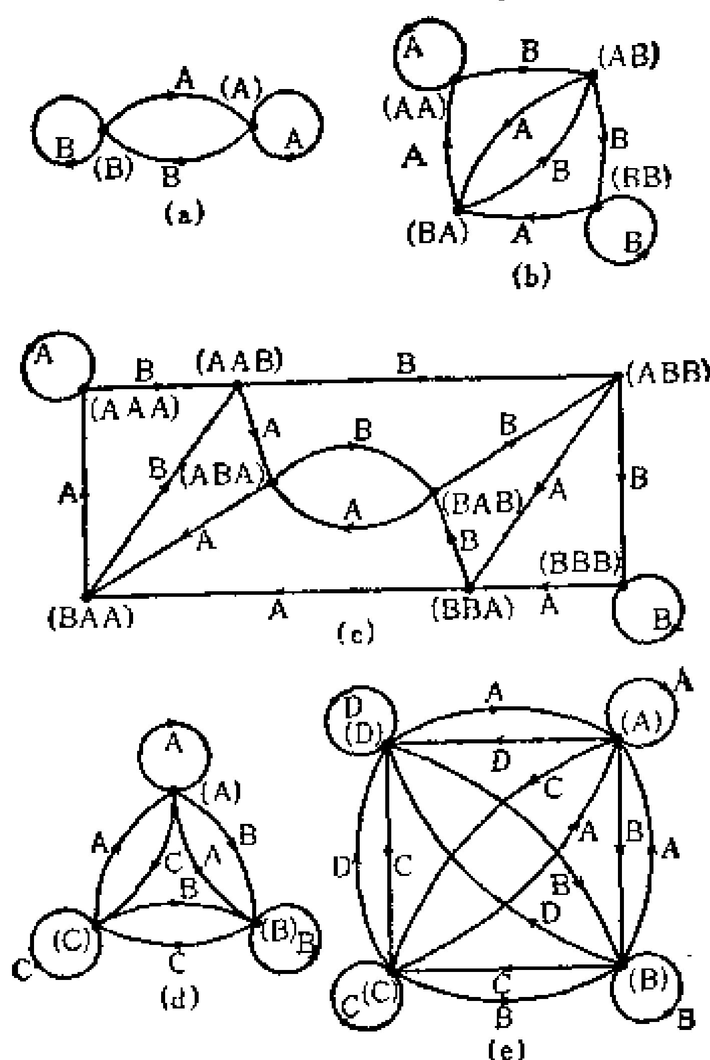


图 7.1

(*a*) 单純 Markoff 过程(A, B 两个字母)； (*b*) 2重 Markoff 过程(A, B两个字母)； (*c*) 3重 Markoff 过程 (A, B 两个字母)； (*d*) 3重 Markoff 过程 (A, B, C 三个字母)； (*e*) 单純 Markoff 过程(A, B, C, D 四个字母)

方向綫，若其迁移概率为零，方向綫实际上不存在。

設由任一状态出发可以达到任何其他状态，这种 Markoff 过程叫**既約 Markoff 过程**[13]，如果从某一状态 E_j 出发，又回到 E_j 的概率为 1，即一定能返回原来的状态时，这种状态 E_j 叫回归状态❶ (recurent state)。反之，若概率小于 1，即不一定能回到原来

❶ 如果回到原来状态所需要的沿方向綫迁移的平均次数为无限时，这种回归状态称为零状态(null state)。因为状态个数为有限时，上述情况不会发生，故本书以后不考虑这种情况。

的状态，这种状态就叫**过渡状态** (transient state)。又 E_j 經过 t, $2t$, $3t$, … (t 为具有如此性质的迁移次数的最大整数，它大于 1) 仍返回状态 E_j 时，則这种状态叫**周期状态** (periodic state)。不是零状态，也不是周期状态的回归状态特别叫做**遍历状态** (ergodic state)。例如图 7.1(b)，由状态 (AB) 沿着字母 B 的方向綫到状态 (BB) 的迁移概率为零，沿其他方向綫的迁移概率为正。于是一旦由状态 (BB) 迁移到状态 (BA)，便不再回到状态 (BB) 了，这个状态 (BB) 就叫过渡状态，此外，其他三个状态一定回到原来状态时叫回归状态，且显然是非周期性的，故此三状态称为遍历状态。又当状态 (AB)，(BA) 間两个方向綫及状态 (AA)，(BB) 联結的閉合綫的迁移概率是零时，其余的方向綫都变成了周期状态。

有限个状态的既約 Markoff 过程，是由一切的周期性的或一切的非周期性的回归状态所組成。因此，一个不是周期性的状态便是遍历状态。設一状态由 E_j 出发，經 n 次迁移后变成 E_k 的迁移概率为 $p_j^{(n)}(k)$，則存在一个满足关系式

$$\lim_{n\to\infty} p_j^{(n)}(k) = u_k > 0 \tag{7.1}$$

的 u_k，此值与 j 无关。即在某一时刻后，状态 E_k 的概率与起始分布无关，达到概率上的平衡。設状态 j 一次迁移到状态 k 的迁移概率为 $p_j(k)$，各个 u_k 对一切状态的和为

$$\sum_k u_k = 1, \tag{7.2}$$

具有上列性质的概率分布满足以下的公式：

$$u_k = \sum_j u_j p_j(k). \tag{7.3}$$

上式右端对一切 j 取和。这种 u_k 称为**平稳过程（絕对）概率**。当状态間的迁移概率 $p_j(k)$ 都已知时，概率分布 u_k 便由 (7.2) 及 (7.3) 两式唯一决定。

下面的信息論只限于考虑遍历的 Markoff 过程。現設状态

$(L_1, L_2, \cdots, L_n)$作为信息源，过去发生的字母序列 $L_1, L_2, \cdots, L_n$ 由上述綫图可以清楚地看出，由此状态到下一个状态发生字母 L_k 的迁移概率，即此信息源在过去发生序列 $L_1, L_2, \cdots, L_n$ 的条件下，接着发生 L_k 的条件概率为 $p_{L_1\cdots L_n}(L_k)$. 就是說，当已知一遍历的信息源的字母間的条件概率 $P_{L_1\cdots L_n}(L_k)$ 时，表示此信息源的綫图的状态間的迁移概率就被确定了，即由(7.2)，(7.3)两式，可求出各个状态間的概率。信息源在状态$(L_1, L_2, \cdots, L_n)$的平稳过程概率不过是在此信息源发生字母序列 $L_1, L_2, \cdots, L_n$ 的联合分布概率 $P(L_1, L_2, \cdots, L_n)$。

为了表示信息源的随机过程性质，Shannon 引入了以上的綫图表示法，实际上很簡便。即在上述表示中，由一个状态到另一个状态的方向綫最多只有一条，在概率的意义下，几个同一的状态可并在一起，即两个状态間不論有多少方向綫，总可以联在一起。这种表示方法比前述的表示信息源的 Markoff 过程性质的方法簡洁，而且同样正确，以后在定义信息源的熵那样重要的量时，与未簡化的綫图有相同的值，这是很有利的。

上述的綫图中把两个状态統一起来，先对所討論的問題中两个状态看一下进入各状态的方向綫，由一状态发出的方向綫进入另一状态（問題中两个状态同样看待），在原来的那一状态必有方向綫进入。其次看問題中由两个状态所发出的綫路，由一个状态发出的方向綫到达另一个状态(問題不只是两个状态，而且把两个状态分别同样看待)，另一状态发出同一迁移概率的方向綫。若有以上两个性质，这两状态便統一为一个状态，又当进入两个状态的方向綫一般說由不同的字母发生，若統一为一个状态时，由另外的状态可画两条方向綫，且各条方向綫是由不同的字母所发生。

例如現在来看由点、划、字母間隔、单語間隔四个字母組成的，且字母間隔与单語間隔以及字母間隔本身及单語間隔本身不能相继出現两次以上的

单純 Markoff 过程 Morse 碼所作的綫图。图 7.1(e) 的 A, B, C, D 各表示点、划、字母間隔以及单語間隔。因为单語間隔、字母間隔分别不能出現两次以上，所以除去状态(C)，(D)这一組，又因为两者不能相继出現，所以省掉图中(C)，(D)間两条方向綫，变成图 7.2(a)的样子。此外，由状态(A)，(B)出发，点、划、字母間隔、单語間隔发生的概率相同，各以 p, q, r, s 表示，(A)除了向状态(B)，(C)，(D)以外，向(A)本身也发出方向綫，(B)也是向(A)，(B)，(C)，(D)发出方向綫。其次由(A)，(B)出发的方向綫都到达(A)，(B)，(C)，(D)，而且有相同的迁移概率，所以可把(A)，(B)統一起来成为一个状态。同样由(C)，(D)出发的点、划的概率設为 a 和 b，显然也可以統一起来，結果变成了图 7.2(b) 的綫图。

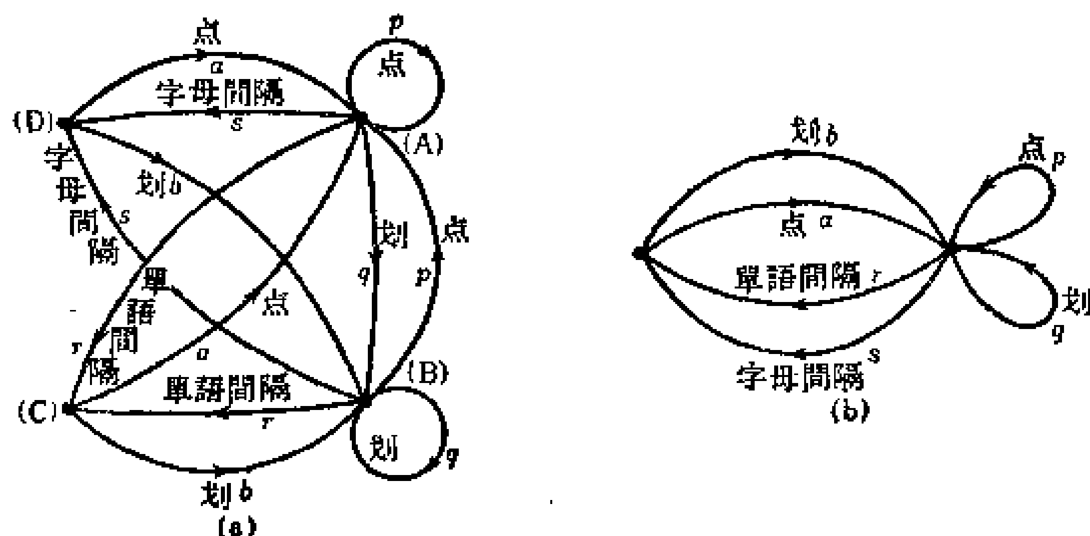


图 7.2

在特殊情况下，如果信息源过去发生的字母与以下发生的字母完全独立，则状态的个数只有一个。即如图 7.3 所示。

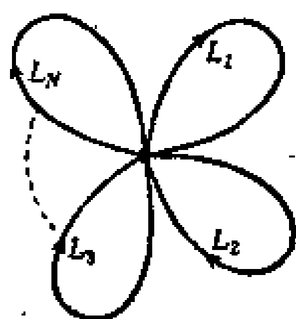


图 7.3

用状态和方向綫統一形式表示信息源的綫图叫 **Shannon 綫图**。去掉迁移概率为零的方向綫的綫图，表示对信息源发生字母的方法有所限制。

以后所述的推广的 Shannon 綫图中計算各状态的平稳过程

概率 u_k 时，設由状态 j 到 k 有 s 条方向綫，其迁移概率各为 $p_j(k,1)$，$p_j(k,2)$，…，$p_j(k,s)$，則(7.3)式中的 $p_j(k)$ 可以換成 $\sum_{t=1}^{s} p_j(k,t)$ 再进行計算。

§8 作为信息源的自然語

人們在交往中，自然发生了語言的使用，但自然語究竟怎样能够看做前节所述的数学的随机过程呢？

今举英文的普通文章为例。字母序列要在文法上正确，且有意义，在这样的文章中，各字母出現的頻率是不会一致的。不只是英文，任何自然語都有相当明显的这种傾向。虽然由于文章的性质不同，例如政治評論和小說，多少有些差异，但大体是一样的。英文 26 个字母加上单語間隔共 27 个字母，其出現頻率經調查后如表 8.1 所示[14]。

但是反过来，26 个字母及单語間隔所成的 27 个字母依照表 8.1 的頻率无次序地排列，得到(英文的第一次近似)：

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHE NHTTPA OOBTTVA NAH BRL

这是用英文 26 个字母排成的象英文但实际上不是英文的排列，因为沒有意义。英文是有意义的，由于字母的排列方法而使之有意义这件事很重要。例如字母 Q 接下去是 U，字母 E 接下去是 D，字母 T 接下去是 H 的概率較大。自然語是有意义的字母排列，因此以前出現过什么样的字母排列是十分重要的事，前节所論，若不考虑过去若干个 (n 个) 字母决定以下出現字母的条件概率，就不能掌握自然語的特点。于是 n 愈大，愈能正确地預測自然語的概率性质。先考虑 $n=1$ 的情况，卽考虑前一个字母，用前节的讲法卽以单純 Markoff 过程来看。如果已經調查清楚文章中出現字母 i 时，其次出現字母 j 的条件概率为 $p_i(j)$，那末反过来可

表 8.1

頻繁次序	字母	頻率	頻繁次序	字母	頻率
1	单詞間隔	0.1817	15	M	0.02075
2	E	0.1073	16	U	0.02010
3	T	0.0856	17	G	0.01633
4	A	0.0668	18	Y	0.01623
5	O	0.0654	19	P	0.01623
6	N	0.0581	20	W	0.01260
7	R	0.0559	21	B	0.01179
8	I	0.0519	22	V	0.00752
9	S	0.0499	23	K	0.00344
10	H	0.04305	24	X	0.00136
11	D	0.03100	25	J	0.00108
12	L	0.02775	26	Q	0.00099
13	F	0.02395	27	Z	0.00063
14	C	0.02260			

以用条件概率把无次序的字母排列成下列样子(第二次字母近似)❶:

ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

其次 $n=2$ 时，即将2重 Markoff 过程的字母随意地排列成(第三次字母近似):

IN NO IST LAT WHEY CRA TICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OFTHE REPTAGIN IS REGO ACTIONA OF CRE

---

❶ 可先用下述方法作成一个随机数表。先将一书任意打开，任意写下一字母，再任意打开一頁，讀下去，若遇到前一次出現的字母时，再写下此字母下面一个字母。然后又任意打开一頁讀下去，发現第二个字母时，又写下此字母下面的一个字母。此法继續下去，即得第2次字母近似。

以下高次近似也用同法处理。

今假設使用单語，按出現的頻率加以随机排列而得（第一次单語近似）：

REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE

若考慮前一单語时（第二次单語近似）：

THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED

經过上述各个阶段，逐漸地就与英文的文章相接近了。但在第二次字母近似时，两个字母序列相当妥当且有意义，但四个字母序列就不常出現在普通的文章中。又在第二次单語近似中，大部分象英文的样子，但四个左右的单語序列就越出正軌，有意义的文章就不多了。但若考慮了 Markoff 过程中較长的过去事件，离散的信息源可以很正确地表示出来。

已知字母的迁移概率时，用 Shannon 綫图的表示方法已在前节中讲过，此处假定已知单語的頻率分布，而用例子說明 Shannon 綫图。設有 A, B, C, D, E 五个字母組成 16 个单語的且頻率分布如下的信息源：

A	0.10	BEBE	0.16	CABED	0.11	DEB	0.04
ADEB	0.04	BED	0.04	CEED	0.05	DEED	0.15
ADEE	0.05	BEED	0.02	DAB	0.08	EAB	0.01
BADD	0.01	CA	0.05	DAD	0.04	EE	0.05

这些单語可按这样的頻率随机地写成如下的文章：

DAB EE A BEBE DEED DEB ADEE ADEE EE DEB BEBE BEBE BEBE ADEE BED DEED DEED CEED ADEE A DEED DEED BEBE CABED BEBE BED DAB DEED ADEB

此种信息源以方向綫表示的对应于各字母的 Shannon 綫图就象图 8.1 所示，其中 S 表示間隔。

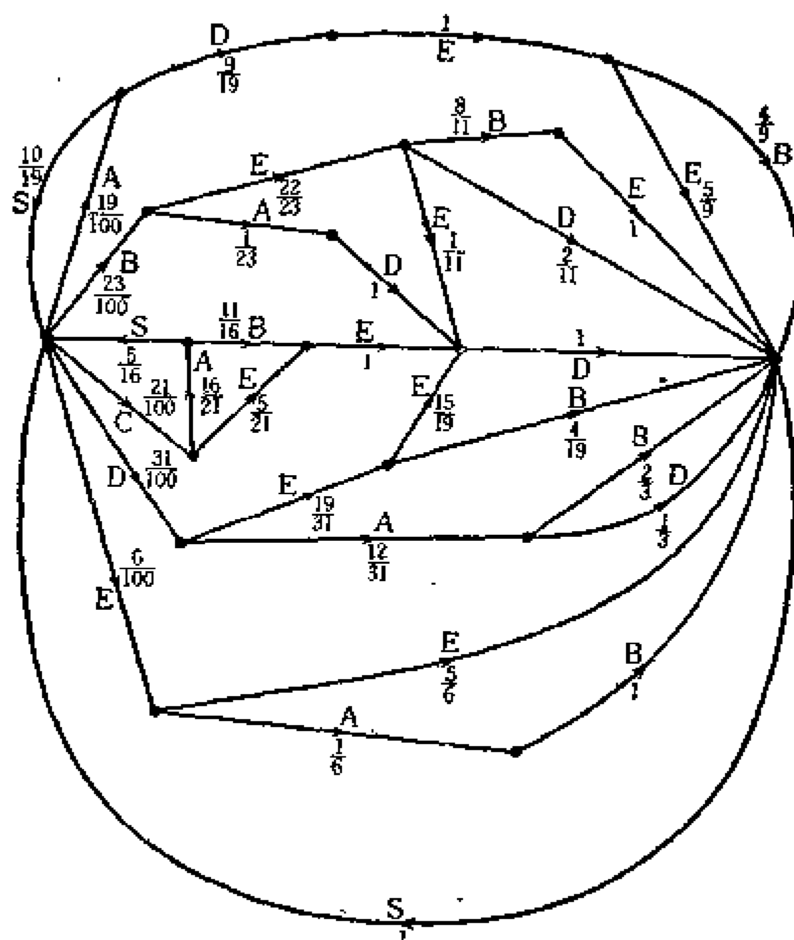


图 8.1

§9 信息源的熵

对于信息源的消息，若过去发生的字母是独立的字母序列，那末图 7.3 表示的信息源的熵已在 §4 内定义过。这里我們将定义 Markoff 过程表示的消息的信息源的熵。

在 §7 中所述信息源只是由有限个遍历状态表示的 Shannon 綫图。其中状态 E_i 到状态 E_j 的第 s 条方向綫沿箭头方向的条件

概率为 $p_i(j,s)$（图9.1）。对于沿第 s 条方向綫由 E_i 到 E_j 的信息源，考虑为产生长度 $l_{ij}^{(s)}$（以秒表示）的文字 $L_{ij}^{(s)}$ 的信息源。

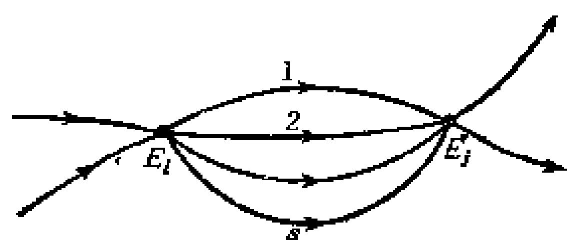


图 9.1

若由状态 E_i 发生字母的一切可能的熵为 H_i，取 E_i 的平稳过程絕对概率 P_i 的加权平均，则可定义**信息源的一个字母的熵**为

$$H=\sum_i P_iH_i=-\sum_i P_i\sum_{j,s}p_i(j,s)\log p_i(j,s)\quad(\text{熵/字母})。\tag{9.1}$$

此即字母平均的熵。設在1秒钟內发生字母的平均数为 m，单位时間的熵即定义为

$$H'=mH\quad(\text{熵/秒})。\tag{9.2}$$

这个 m 有下列的关系式

$$m=[\sum_{i,j,s}P_ip_i(j,s)l_{ij}^{(s)}]^{-1}.\tag{9.3}$$

(9.1) 式的定义是 §7 最后所述的状态間合一的 Shannon 綫图的推广形式，但是，在 §7 的前半所述的狭义綫图中，S 最多只是1，由状态、方向綫的合一，很容易証明同样的熵是与(9.1)式的熵一致的。

用字母单位表示的熵 H 与用时間单位表示的熵 H' 之間的区别很重要。前者沒有时間的概念，也可以說是信息源內靜止的內蘊的信息量。例如在一本书中一个字母平均含有的信息量。現在我們讀这本书时，信息随着时間由口中的声音发出来。后者是在单位时間由信息源发生的信息量。

以上，由信息源各字母出現的概率定义了它的熵，現在所談的不是字母本身，而是由几个字母相联构成的序列，亦即从消息发生的概率来求信息源的熵。为了簡单起見，設消息中的字母与过去发生的字母是独立的。設字母 L_i 的概率为 p_i，信息源的熵如 §4

所述为$-\sum_i p_i \log p_i$. 如果考虑相当长的n个字母所作成的消息，則由大数法則，字母L_1出現$p_1 n$个，字母L_2出現$p_2 n$个，… 的概率是很高的，随着n的增大，其概率逐渐接近于1. 从而含有字母$L_1, L_2, \cdots, L_N$的特定消息出現的概率約为

$$p = p_1^{p_1 n} p_2^{p_2 n} \cdots p_N^{p_N n}. \tag{9.4}$$

取对数得

$$\log p = n \sum_{i=1}^{N} p_i \log p_i.$$

所以 $\log p \approx -nH$, 由此得出

$$H \approx \frac{\log \frac{1}{p}}{n}. \tag{9.5}$$

所以信息源的熵近似地等于其所发生的典型长度消息的概率倒数的对数，除以字母的个数。

上述消息恰好是n个字母的字母序列时，这个消息的长度叫n个字母。

实际上，(9.5)是一个近似式，很容易証明它对遍历过程所表示的信息源也是成立的。这个 Shannon 定理經国澤严密証明了，并且求出消息的概率分布[15]，本多也曾論証过这个問題[16]。

定理 9.1 設长度为n个字母的消息的发生概率为p，当n充分大时，$-\log p$的概率分布，对于任意实数$a<b$，近似地等于

$$P\{a < -\log p < b\} \approx \frac{1}{\sqrt{2n\pi}\,\sigma_H} \int_a^b \exp\left[\frac{-(x-nH)^2}{2n\sigma_H^2}\right] dx, \tag{9.6}$$

式中H为信息源对单位字母的熵，又

$$\sigma_H^2 = \sum_{i,j,s} P_i p_i(j,s) [\log p_i(j,s)]^2 - H^2$$
$$+2 \sum_{i,j,s} P_i p_i(j,s) \log p_i(j,s) \sum_{k,l,m} \Omega_{jk} p_k(l,m) \log p_k(l,m), \tag{9.7}$$

$$\Omega_{jk} \equiv \lim_{n\to\infty} \sum_{t=0}^{n-1} [p_j^{(t)}(k) - P_k] \tag{9.8}$$

此处 $p_j^{(t)}(k)$ 为由状态 E_j 經 t 条方向綫，即 t 个字母，到达状态 E_k 的概率（上式中的对数或取自然对数，或取以 2 为底的对数，但必須取得一致）。

由上列定理立即知道，当 n 充分大时，$-\log p$ 的概率分布是以 nH 为中心的非常急峻的正态分布。給定任意小的正数 ε 和 δ，对于充分大的 n，由此信息源所产生的消息可分成两組。消息概率之和比 ε 小的为第一組，其他滿足公式

$$\left|\frac{\log p^{-1}}{n}-H\right|<\delta$$

概率为 p 的消息为第二組。

各字母的时間长度不相同时，若以 n 个字母长度的消息的时間长度代替 $-\log p$；則上述定理 9.1 就成为关于由信息源发出的消息时間长度的概率分布定理。但 H 須取字母的平均时間长度 (9.3)，σ_H^2 取为时間的偏差。

§10 信息源的冗长度

熵这一概念不仅对于电通信系統的信息源很重要，而且对于英文、德文、日文等自然語的語言結构研究也是很便利的概念。

已知某一信息源时，設其熵为 H。再假想另一个有相同字母的，且这些字母是与过去字母独立发生的信息源，并設各字母发生的概率都相同，其熵为 $H_{\max}$，則

$$H/H_{\max} \tag{10.1}$$

叫做**相对熵**(relative entropy)。这个量用普通的百分比表示，其物理意义正如以后会談到的，就是这两个信息源发出的消息变成相同的代碼时，給定的信息源比第二个信息源所压縮的程度。由 1 减去相对熵，即

$$1-(H/H_{\max}) \tag{10.2}$$

叫做**冗长度**(redundancy)。它表示不同种类的字母在一定的情况下，这个給定的信息源含有多少沒有用的东西。

例如某种語言的冗长度为 80%，我們将此語言写成文字时，有 80% 是由語言結构规定好的，剩下的 20% 是写的人自由选擇的。

日本語的冗长度只用平假名❶和汉字写的文章是两个极端的情况。又如英語的"基础英語"与 James Joyce 所著"Finnegans Wake"也是很好的例子。"基础英語"的語汇，限于用 850 个单語，其冗长度极高。例如"眉"不用一个单語，而用"眼上面排列的毛"这样复杂的表达。与此相反，Joyce 的語汇非常丰富，許多意义压縮在一个单語内，冗长度很小。

下面一个关于上述自然語的复杂信息源的定理，不假設它作为 Markoff 过程的状态及状态之間的迁移概率，而是从消息表現的頻率中直接計算其熵，这是很方便的。

定理 10.1 設信息源发出长度为 n 个字母的消息 B_i 的概率为 $p(B_i)$，在 B_i 之后发生字母 S_j 的概率为 $p(B_i, S_j)$，而 $p_{B_i}(S_j) = p(B_i, S_j)/p(B_i)$ 为出現 B_i 以后，又发生 S_j 的条件概率。令

$$G_n = -\frac{1}{n}\sum_i p(B_i)\log p(B_i), \tag{10.3}$$

$$F_n = -\sum_{i,j} p(B_i, S_j)\log p_{B_i}(S_j), \tag{10.4}$$

第一式的 $\sum$ 是长度为 n 个字母一切系列 B_i 的总和，第二式的 $\sum$ 表示长度为 $n-1$ 个字母的一切 B_i 以及 S_j 一切字母的总和。則 G_n, F_n 为 n 的单調递減函数，同时，

$$F_n = nG_n - (n-1)G_{n-1}, \tag{10.5}$$

$$G_n = \frac{1}{n}\sum_{k=1}^{n} F_k, \tag{10.6}$$

---

❶ 平假名是日文一种字母，始于平安时代，是一种草书体。——譯者注

$$F_n \leqslant G_n \tag{10.7}$$

等关系成立，且

$$\lim_{n\to\infty} F_n = \lim_{n\to\infty} G_n = H. \tag{10.8}$$

証明的要点如下：首先，F_n 为 n 的单调递减函数是很清楚的，在

$$\begin{aligned} F_n &= -\sum_{i,j} p(B_i, S_j)\log p(B_i, S_j) + \sum_{i,j} p(B_i, S_j)\log p(B_i) \\ &= -\sum_{i,j} p(B_i, S_j)\log p(B_i, S_j) + \sum_{i} p(B_i)\log p(B_i) \end{aligned}$$

中代入 G_n 的定义，就得到(10.5)式。此式对一切 n 取和便得(10.6)，于是得(10.7)式，G_n 显然也是单調递减函数。并且 F_n 及 G_n 必須收斂于同一极限値。利用定理9.1，得 $\lim_{n\to\infty} G_n = H$，由(10.7)式得 $\lim_{n\to\infty} F_n = H$。

这个定理中的 F_n 为已知过去 $n-1$ 个字母时对于下一个字母的条件熵，G_n 为 n 个字母序列的熵。这个定理研究了长度为 n 个字母的消息的統計构造，当 n 逐漸增大时，F_n 逐漸接近于正的 H 是显然的。实际上，F_n 如§8所述是自然語第 n 次的近似熵。在信息源为 r 重 Markoff 过程的特殊情况下，对于一切的 $n>r$，都有 $F_n = H$。

关于26个字母，各国語言的条件熵 F_n 經实际調查的結果如下表所示[17]。

	英語	法語	德語	西班牙語
F_0	4.70 必特	4.7 必特	4.7 必特	4.70 必特
F_1	4.124 必特	3.984 必特	4.095 必特	4.015 必特
F_2	3.56❶必特			
F_3	3.3❶ 必特			
F_w	1.65 必特	3.02 必特	1.08 必特	1.97 必特

一个单語所含的平均字母数为：英語4.5个字母，法語4.8个字母，德語5.92个字母，西班牙語为4.96个字母。此表是以原体

❶ Shannon[18] 的結果。

字为基础作成的，上述定理的 $\bar{F}_n$ 是由单語的频率表算出。F_w 是由单語的频率求得的熵，再除以单語所含的平均字母个数，Shannon不用原体字而用不同的单語频率表計算英語的 F_w 为 2.62 必特[18]。由上表看出，德文单語的平均长度最长，但其熵最小。

再談一下英語，由此表得出冗长度大約为 65%，Shannon[18] 曾考查过大約过去 100 个字母影响下英語的熵，一个字母为一必特左右的值，此冗长度为 75%。换言之，写英文时，有 25% 是个人自由选擇的字母。若 100 个字母以上过去都考虑在内时，实测的冗长度就更大了，此时由于包括了文体与内容的影响，所以可以得到可靠的結果。

第4章 无噪音的离散信道

本章首先叙述作为发射机及接收机的本质的变换作用，其次讲述沒有噪音干扰的离散信道，导出表示信道信息傳送能力的通信容量概念，阐述有关的基本定理，并論述能得到最佳的信息傳送效率的編碼方法。

§11 作为变换器的发射机及接收机的功能

Morse 电报信号是由点、划、字母間隔、单語間隔四种不同时間长度的代碼組合而成的代碼序列。电傳打字电报的信号是由64种不同时間长度的代碼組合而成的序列。所謂离散的信道，乃是一个傳輸如 Morse 电报、电傳打字电报那样以有限种类的代碼組合而成的代碼序列的信道，亦即考虑离散的变数。

首先，发射机将信息源的字母序列变成适当的代碼序列信号送入信道。例如对于 Morse 电报的情况，电文先变成点和划的組合发出去。接收机則把它还原。若着眼于这种变换作用，发射机和接收机都是一种**变换器**(transducer)。这种变换器的一般性质考虑如下。

一般地說，变换器的內部有記忆的作用，其輸出的代碼不仅与現在輸入的字母有关，而且与过去的字母有关。假定內部的記忆是有限的，即变换器可以取 m 个不同状态，当信息源的字母进入变换器时，m 个状态在內部发生迁移，此时假設变换器輸出的代碼唯一地由內部状态和輸入字母所決定。且那个內部状态的迁移，由現在的內部状态和輸入字母唯一地决定。即設 x_t 为 t 时刻輸入的字母(可能是字母序列)，令 β_t 为此时变换器內部的状态

(β_t 为 $\beta_1, \cdots, \beta_m$ 中之一)，y_t 为当变换器内部状态为 β_t 及输入字母 x_t 时的输出代码(可以是代码序列)，则

$$y_t = f(x_t, \beta_t), \tag{11.1}$$

$$\beta_{t+1} = g(x_t, \beta_t). \tag{11.2}$$

若将一个变换器的输出码看成为第二个变换器的输入，则此两个链索式联接的变换器可以看成为一个变换器。在此情况下，如果第二个变换器的输出完全等于第一个变换器的输入字母，则第一个变换器叫做**正规**的，第二个变换器叫做它的**反变换器**。通常发射机是正规的变换器，接收机为其反变换器。

信息源联接变换器时，其输出仍可考虑为信息源，输出的代码序列可以当作等价信息源的消息，于是变换器输出的代码序列的熵，可以按§9所述的方法計算。这个熵与原来信息源的熵之間的关系如下。

定理 11.1 与某信息源联接的具有有限个内部状态的变换器的输出，可以考虑为一个信息源，其熵比原来的信息源的熵小或相等。如变换器为正规的，则两者相等。

証明 假定信息源的状态为 α，变换器内部的状态为 β，并設信息源由状态 α_1 迁移到 α_2 时发生字母 x_1. 与信息源联接的变换器的输出，可考虑成状态为 α 和 β 所表示的状态的积 (α, β) 的第二个信息源。信息源的状态为 α_1，变换器内部的状态为 β_1，信息源由 α_1 变到 α_2 产生字母 x_1 时，如果变换器由 β_1 变到 β_2 产生代码 y_1，则可以把两个积状态 (α_1, β_1)，(α_2, β_2) 用箭头联接，且原来信息源 α_1 产生 x_1 的迁移概率可用 $p_{\alpha_1}(x_1)$ 表示。第二个信息源的熵于是能由这种积状态及由联接的方向綫所成的 Shannon 綫图計算出来。从原来的信息源状态 α_1 迁移，如果有产生字母 $x_1, x_2, \cdots, x_k$ 的可能性，那末这个状态所对应的熵为

$$-P_{\alpha_1} \sum_{j=1}^{k} p_{\alpha_1}(x_j) \log p_{\alpha_1}(x_j), \tag{11.3}$$

其中 P_{α_1} 表示取状态 α_1 的概率。取遍一切状态的和，就可以得出信息源的熵。这对于第二个信息源說，它的积状态含有 α_1 者設为 $(\alpha_1, \beta_1), \cdots, (\alpha_1, \beta_l)$

設占有各状态的概率为 $P_{\alpha_1\beta_1}, \cdots, P_{\alpha_1\beta_l}$，則有下列关系：

$$P_{\alpha_1}=\sum_{h=1}^{l} P_{\alpha_1\beta_h}. \tag{11.4}$$

从而由这些状态求出熵为

$$-\left[P_{\alpha_1\beta_1}\{p_{\alpha_1}(x_1)\log p_{\alpha_1}(x_1)+\cdots+p_{\alpha_1}(x_s)\log p_{\alpha_1}(x_s)\}+\cdots\right.$$
$$\left.+P_{\alpha_1\beta_l}\{p_{\alpha_1}(x_u)\log p_{\alpha_1}(x_u)+\cdots\}\right]. \tag{11.5}$$

只有在这些状态中由任一状态一致产生字母 $x_1, x_2, \cdots, x_k$ 时才与(11.3)式相等。即(11.5)中任一{ }中的式子等于式子 $\sum_{j=1}^{k} p_{\alpha_1}(x_j)\log p_{\alpha_1}(x_j)$ 时，利用(11.4)可以得出(11.5)与(11.3)相等。但在相反的情况下，只发生 $x_1, x_2, \cdots, x_k$ 的一部分时，这个式子比(11.3)的数值小。从而由全体来看第二个信息源的熵并不增大。定理的后半証明如下：正規变换器的輸出恰好联接其反变换器。若以 H_1', H_2' 和 H_3' 分别表示信息源的熵、正規变换器輸出的熵和反变换輸出的熵，則 $H_1' \geqslant H_2' \geqslant H_3' = H_1'$，所以 $H_1' = H_2'$ 成立。

以下所論的发射机及接收机假定是正規的变换器及反变换器。特别对于无噪音离散信道，联接的信息源和发射机是正規变换器时，其輸出的代碼序列不会受到干扰，接收机是反变换器。从而在这种情况下，信道簡单地只是正規变换器及反变换器之間的具有信息变换作用的物体。本章以下將討論这种变换作用。

§12 无噪音离散信道的傳輸速率和通信容量

在无噪音信道中傳輸信息时，如何确切地定义傳輸能力呢？对于这种信道，如果傳輸信息源的信息量为 H，由于沒有噪音的干扰，信息量沒有損失，所以送到接收机中的信息量仍为 H。

定义 12.1 已知某一无噪音离散信道联接着某一信息源，当它发出 H' (熵/秒) 的信息量时，称 H' 为这个信道的**傳輸速率** (transmission rate)，与此信道可能联接的一切信息源中，傳輸速率 H' 的最大值叫此信道的**通信容量** (capacity)，以 C 表之。

以上是以熵为基础来定义的通信容量，与此完全不同，也可由实际的物理观点来定义，即某一信道的通信容量可定义为傳輸速

率的最大值：

定义 12.2　設給定一个无噪音信道，在时間 T 之間，此信道上傳輸不同的信号（即信号序列）的总数为 $N(T)$，則通信容量 C 定义为

$$C=\lim_{T\to\infty}\frac{\log N(T)}{T}. \tag{12.1}$$

現在來証明通信容量的定义 12.2 与定义 12.1 是完全等价的。

首先解决最簡単的情况，把信道中的代碼序列看做由信息源发出的消息，假設出現的代碼完全与过去的代碼独立，也就是与 Morse 电报等情况不同，不論什么代碼都允許出現。在此情况下，若一秒时間內有 M 个代碼，等价信息源的熵为 H'，則在一切代碼有相同的概率时，熵的值达到最大，其值为 $\log M$。根据定义 12.1，C 即为 $\log M$，但由(12.1)式因 $N(T)=M^T$，故 $C=\log M$，从而这两个通信容量是一致的。

在更一般的情况，可以考虑此等价的信息源为一 Markoff 过程，而且是遍历过程，也就是信道上的代碼序列附加了这样或那样的限制的情况。使用 Shannon 綫图表示时，設由状态 E_i 到 E_j 第 s 个方向綫的迁移概率为 $p_i(j,s)$，产生长度为 $l_{ij}^{(s)}$ 的代碼，状态 E_i 的概率为 P_i，由(9.1)和(9.3)式以及定义 12.1，傳輸速率为

$$H'=\frac{-\sum_{ijs}P_ip_i(j,s)\log p_i(j,s)}{\sum_{ijs}P_ip_i(j,s)l_{ij}^{(s)}}\ (\text{熵/秒}). \tag{12.2}$$

并且，因为是遍历过程，所以由(7.3)的关系有

$$\sum_i P_i\sum_s p_i(j,s)=P_j \tag{12.3}$$

及

$$\sum_{js}p_i(j,s)=1, \tag{12.4}$$

$$\sum_i P_i=1. \tag{12.5}$$

若 $p_i(j, s)$ 已經确定，由上式也决定了 P_i. 从而求(12.2)的最大值，只要求(12.3)乃至(12.5)条件下的 $p_i(j, s)$ 的最大值。求(12.2)的极大値，可用 Lagrange 的待定乘数法，它是表明在概率非負的范圍內成为最大值。不管条件是什么，考虑 $p_i(j, s)$ 的极大，只需考虑下式中关于 $p_i(j, s)$ 及 P_i 的极大值就够了。

$$U = \frac{-\sum_{ijs} P_i p_i(j, s) \log p_i(j, s)}{\sum_{ijs} P_i p_i(j, s) l_{ij}^{(s)}} + \sum_j \lambda_j \left(\sum_{is} P_i p_i(j, s) - P_j \right)$$

$$+ \sum_i \mu_i \left(\sum_{js} p_i(j, s) - 1 \right) + \gamma \left(\sum_i P_i - 1 \right). \qquad (12.6)$$

式中 λ_j, μ_i, γ 是待定乘数。所以，

$$\begin{cases} \dfrac{\partial U}{\partial p_i(j, s)} = \dfrac{-P_i\{1 + \log p_i(j, s)\}}{m} - \dfrac{P_i l_{ij}^{(s)} \cdot C}{m} + \lambda_j P_i + \mu_i \\ \qquad = 0, & (12.7) \\ \dfrac{\partial U}{\partial P_i} = \dfrac{-\sum_{js} p_i(j, s) \log p_i(j, s)}{m} - \dfrac{C \sum_{js} p_i(j, s) l_{ij}^{(s)}}{m} \\ \qquad + \sum_{js} \lambda_j p_i(j, s) - \lambda_i + \gamma = 0, & (12.8) \end{cases}$$

上式中(12.2)的极大值用 C 表示，其分母碼的平均长度用(9.3)式的 m 代替。(12.8)式中乘以 P_i 再取一切 i 之和得 $\gamma = 0$. (12.7)式乘以 $p_i(j, s)$，取一切 j, s 之和，用此和式与(12.8)得到 μ_i 的关系式，再代入(12.7)，于是得到 $\log p_i(j, s)$ 的表示式，对数还原之后，得

$$p_i(j, s) = \exp(-C l_{ij}^{(s)} + m\lambda_j - m\lambda_i). \qquad (12.9)$$

取一切 j, s 之和再利用(12.4)式，由上式又可得

$$\sum_{js} \exp(-C l_{ij}^{(s)} + m\lambda_j) = \exp(m\lambda_i). \qquad (12.10)$$

这是 $\exp(m\lambda_i)$ 的齐次方程，为了存在非零解，其系数行列式必須等于零：

$$\left| \sum_s \exp(-C l_{ij}^{(s)}) - \delta_{ij} \right| = 0. \qquad (12.11)$$

式中 δ_{ij} 当 $i=j$ 时取值为1，当 $i\neq j$ 时取值为零。从而 C 是(12.11)式的最大正实根。这时迁移概率 $p_i(j,s)$ 由(12.9)給出，是非負的。P_i 可以由(12.3)及(12.5)求出。所以 C 即为所求的通信容量。

其次由定义12.2出发求通信容量。到状态 E_i 为止，設有时間长度 T 的代碼序列数目 $N_i(T)$．由状态 E_i 轉移到 E_j 时，产生了长度为 $l_{ij}^{(1)}, l_{ij}^{(2)}, \cdots, l_{ij}^{(s)}$ 的代碼，从而

$$N_j(T)=\sum_{is} N_i(T-l_{ij}^{(s)}). \qquad (12.12)$$

此为綫性差分方程，当 $T\to\infty$ 时其漸近解是

$$N_j(T)=A_jW^T \quad (T\to\infty), \qquad (12.13)$$

将此式代入(12.12)，得

$$A_jW^T=\sum_{is} A_iW^{T-l_{ij}^{(s)}},$$

即

$$\sum_i\left(\sum_s W^{-l_{ij}^{(s)}}-\delta_{ij}\right)A_i=0.$$

于是 A_i 有非零解的条件必須为

$$\left|\sum_s W^{-l_{ij}^{(s)}}-\delta_{ij}\right|=0. \qquad (12.14)$$

这里最大正实根决定了 W。利用这个出现于(12.13)的 W 和定义12.1，即得通信容量 C 为

$$C=\lim_{T\to\infty}\frac{\log\sum A_jW^T}{T}=\log W. \qquad (12.15)$$

这表示行列式(12.11)及(12.14)完全是相同的，所以由任一个定义出发都一样。上述的結果可以总結成下列定理。

定理 12.1 不受噪音限制的代碼序列作成的离散信道中的代碼序列以 Shannon 綫图表示时，由状态 E_i 轉移到 E_j 所发出的 s 个代碼的长度設为 $l_{ij}^{(1)}, l_{ij}^{(2)}, \cdots, l_{ij}^{(s)}$，則这个信道的通信容量 C 为 $\log W$．其中 W 为

$$\left|\sum_{s} W^{-l_{ij}^{(s)}}-\delta_{ij}\right|=0 \tag{12.16}$$

的最大正实根。此时状态間的迁移概率 $p_i(j, s)$ 为

$$p_i(j, s)=\frac{B_j}{B_i} W^{-l_{ij}^{(s)}}, \tag{12.17}$$

式中 B_i 满足下列关系式:

$$B_i=\sum_{sj} B_j W^{-l_{ij}^{(s)}}. \tag{12.18}$$

§ 13 編碼基本定理

电通信中实际上使用的离散信道，傳送的是各种时間长度的很长的代碼序列。对于这种信道，由信息源发出消息时，可能用什么样的平均速率傳送信息呢？下列定理就回答这一問题。由 § 9 的定理 9.1，当含在信息源的消息中的字母非常多的时候，几乎所有的字母序列都以大致相同的概率出現。当时間很长时，信道上的信号方面，几乎所有的代碼序列也是以大致相同的概率出現，以后者的时間长度除前者字母的个数，可以得到单位时間发出的平均字母个数，从这一点出发也可想象有下列的定理。

定理 13.1 設給定的信息源的信息量为 H (熵/字母)，无噪音离散信道的通信容量为 C (熵/秒)，对信息源发出的消息进行适当的編碼，则在此信道上，可能以单位时間平均 $\left(\frac{C}{H}-\varepsilon\right)$ 个字母的速率进行傳输，其中 ε 为任意小正数。但不可能以比平均 C/H 个字母还大的速率进行傳输。

証明 首先，利用定理 11.1，很容易証明这个定理的后一半。以定理中所述的速率傳送信息时，这个发射机內的代碼在单位时間所出現的熵 H' 不会超过通信容量 C，即 $H' \leqslant C$。以信息源的信息量 H 除两端得单位时間的平均字母数 $H'/H \leqslant C/H$，这就証明了定理的后一半。

定理的前半，可以用定理 9.1 所述的性质来証明，今以 **Shannon** 編碼法

具体地說明編碼的方法。长度为 n 个字母的消息按其概率的大小顺序排列，得 $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_M$。其次，取其累加概率 $P_s = \sum_{1}^{s-1} p_i$（其中不含 p_s），将此 P_s 按二进位小数展开，取第一位到第 m_s 位数。这个 m_s 是一个整数，滿足关系式

$$\log_2 \frac{1}{p_s} \leqslant m_s < 1 + \log_2 \frac{1}{p_s}. \tag{13.1}$$

第 s 个消息，就用这样二进位数 0 与 1 得到的二元碼来表示。概率大的消息可用短的二元碼表示，概率小的消息可用长的二元碼表示。由(13.1)得

$$\frac{1}{2^{m_s}} \leqslant p_s < \frac{1}{2^{m_s-1}}. \tag{13.2}$$

P_s 所对应的二元碼与其接下去的 $P_i(i>s)$ 相比較，m_s 位数中至少有一位以上的数字是不同的。因为所有的 $P_i(i>s)$ 对应的二元碼比 P_s 所对应的二元碼至少大于 $1/2^{m_s}$，这个二进位数（注意是小数）展开之后其前 m_s 位都与 P_s 的前 m_s 位不同。从而如此作出的二元碼互相不同，这使消息还元的逆操作終成为可能。其次計算一下原来消息中一个字母所用的二进位碼的平均位数 D_n，得

$$D_n = \frac{1}{n} \sum m_s p_s. \tag{13.3}$$

但由(13.1)及(13.2)，

$$\frac{1}{n} \sum \left(\log_2 \frac{1}{p_s}\right) p_s \leqslant \frac{1}{n} \sum m_s p_s < \frac{1}{n} \sum \left(1 + \log_2 \frac{1}{p_s}\right) p_s, \tag{13.4}$$

从而由定理 10.1 的 G_n 的定义，得

$$G_n \leqslant D_n < G_n + \frac{1}{n}, \tag{13.5}$$

当 n 充分大时，由定理 10.1 知 G_n 接近于信息源以必特/字母所表示的熵 H，再由上式，得 D_n 接近于 H（若信息源的熵的对数以任意数 d 为底时，上述二元碼可换为 d 元碼）。信道中使用二元碼时，就可用上述方法作出的代碼发出去。若在另外的情况，由定义 12.2 知道不管对代碼序列有沒有限制，时間 T 充分大时，在 T 时間內发出去的不同代碼序列的个数大致等于 2^{CT}。这 2^{CT} 个代碼序列可以用 CT 位（实际上取与 CT 相近的整数）的二进位数表示出来，用上述的由 P_s 得到的二元碼按 CT 位分成区間，然后换成此信道固有的代碼序列发出去。从而用此种編碼法在单位时間內平均发出 $CT \dfrac{n}{\sum m_s p_s} \Big/ T$ 个字母，即 C/D_n 个字母。由(13.5)，当 N，T 相当大时，其值接近于 C/H。

一般由发电机供給最大功率至負載时，需要使用变压器，由負載端看来，要求发电机的內部阻抗与負載阻抗相等。这个定理中的代碼变換也有同样的意义，信息源及信道之間可以說进行了統計上的匹配。換言之，由信道側通过变換器所見到的信息源，与信道中可能用最大速率傳送的信息源具有相同的統計性质。完全匹配一般是不可能的，上述定理証明，近似可以达到任意的精度。

Shannon 的編碼法，在发射机及接收机內至少要有 n 个字母的延迟，一般需要很长的延迟时間，这种犧牲以各序列的发生概率有如此长的調节作为补偿。实际上，发射机和接收机延迟了很长时間会使通信变成不可能，所以这种方法沒有太大的实用性。

信道的实际傳輸速率与通信容量之比叫**編碼效率**，上述定理中，此值接近于 1.

§14 最佳編碼法

定理 13.1 是以等时間长度的“0”及“1”二元碼对任意信息源的消息进行編碼，消息中 1 个字母所对应的二元碼的平均位数，与原来消息中的信息量相近，这即是所述的 Shannon 方法。此种二元碼作为实际的信道上的代碼使用起来有很多的便利，而且消息較长时，信息的傳輸效率很好。具有此种性质的編碼方法，还有 Fano 方法[19]，等长編碼法[20]等，这些方法，在消息的长度为有限的情况下都不能保証給出最好的速率。其后，Huffman 发表了有限长度的达到最大平均速率的編碼方法，使这个問題得到了解决[21]。自然，当消息的长度很大时，也收斂于定理 13.1 的理想速率。本节首先叙述容易直观理解的 Fano 編碼法，并与 Shannon 編碼法作一比較，其次說明 Huffman 方法。这里所論的代碼不只是二元碼，也可以推广到多元碼[22]。

Fano 編碼法[23] 首先把 M 个消息按其概率大小順序排列，

其次把此集合分成两个子集，使各子集内的概率之和尽可能接近相等。在此情况下，一个子集内的消息赋予二元碼的最初数“0”，另一个集合的消息赋予“1”．接着把两个子集再分别分解成两子集，第二次子集内的消息概率之和也尽可能使之接近相等，对于第二次分解的集合的消息赋予二元碼的第2数“0”，对于另一个第二次子集内的消息赋予第2数“1”．这个操作继續下去，直到各个子集，只含一个消息为止，使各消息都赋予了一个二元碼。表14.1就表示上述的操作方法，表中的信息源由7个具有所标示概率的消息組成。这个信息源的熵为2.61（必特/消息）。第s个消息所赋予的二元碼的长度設为L_s（秒），碼的平均长度（秒/消息）即是

$$L_{av}=\sum_{s=1}^{M}p_s L_s. \tag{14.1}$$

表14.1　Fano編碼法

消息号数 s	消息概率 p_s	第1次分解	第2次分解	第3次分解	第4次分解	賦予的二元碼	二元碼的长度 L_s
1	0.20	0（概率和0.57）	0 (0.20)			00	2
2	0.19		1 (0.37)	0 (0.19)		010	3
3	0.18			1 (0.18)		011	3
4	0.17	1（概率和0.43）	0 (0.17)			10	2
5	0.15		1 (0.26)	0 (0.15)		110	3
6	0.10			1 (0.11)	0 (0.10)	1110	4
7	0.01				1 (0.01)	1111	4
						平均碼长 L_{av}	2.74

$$傳輸速率=\frac{2.61}{2.74}=0.953(必特/秒)$$

同一信息源，如用定理 13.1 所述的 Shannon 編碼法，則得表 14.2。

表 14.2 Shannon 編碼法

消息号数 s	消息概率 p_s	累积概率 P_s	位数 m_s	賦予的二元碼
1	0.20	0	3	000
2	0.19	0.20	3	001
3	0.18	0.39	3	011
4	0.17	0.57	3	100
5	0.15	0.74	3	101
6	0.10	0.89	4	1110
7	0.01	0.99	7	1111101
		平均碼长 L_{av}	3.14	

$$\text{傳輸速率}=\frac{2.61}{3.14}=0.831\text{(必特/秒)}$$

Shannon 方法及 Fano 方法当消息长度增大时虽說收斂于同一理想值，但在上例有限长度的情况下并不一致。在不一致的时候，Fano 方法的效率比較好[24]。Shannon 方法从数学上說，唯一地确定了二元碼，但 Fano 方法因为对消息集合进行分解，所以不是唯一的。

其次叙述 **Huffman 編碼法**。首先对給定的 M 个消息选取最好的碼組，亦即說明(14.1)的平均碼长最小的碼組要具备五个条件，并闡明滿足这些条件的必然导出最好的碼組。先考虑 d 元碼。

下列两条件是基本的。

(i) 选取的最优碼組中，任取两个碼，具有不同的数碼的排列。

(ii) 当最优碼組的一个长序列已給定时，这个序列最初的碼知道后，其后碼之間的間隔不必标明出来。

例如，取 01，102，111，202 为基本碼，則碼序列：11110220201 可以区分成四个碼 111-102-202-01，但若用 11，111，102，02 作基本碼，就不能作上述碼序列的区分。其次看一下表 14.2 的 Shannon 編碼法，M 个消息按其顺序大小排列，并使概率大的消息对应的碼长比概率小的消息的碼长为小。換成这样的碼，与任意排列时相比，L_{av} 就变小了。換言之，以下列顺序排列：

$$p_1 \geqslant p_2 \geqslant \cdots \geqslant p_{M-1} \geqslant p_M, \tag{14.2}$$

則必得

$$L_1 \leqslant L_2 \leqslant \cdots \leqslant L_M. \tag{14.3}$$

第 M 号消息所对应的代碼的最初 L_{M-1} 位数字，若与第 $M-1$ 号消息代碼的 L_{M-1} 位数字相同，則条件 (ii) 不能滿足，故必不相同，从而把已經有过的碼再加添数字时其位数超过 L_{M-1} 是不必要的。也就是

(iii) $$L_1 \leqslant L_2 \leqslant \cdots \leqslant L_{M-1} = L_M. \tag{14.4}$$

如果 L_M 位数字所构成的消息中，任何两个代碼的前 L_M-1 位数字相异，則把各消息代碼最后一位数去掉之后，可使 L_{av} 变小，而且各代碼間还可以識別。从而得

(iv) 由 L_M 位数字組成的代碼至少有两个，甚至有 d 个代碼，它們的前 L_M-1 位数字相同，而最后一位数字不同。

其次，把 d 种数字排列成 L_M 位以下，假定不作碼来使用，而作为其他碼的开始部分来使用是不可以的。此时所得的数字排列作成的碼能使 L_{av} 变小。亦卽

(v) L_M-1 位所有的数字排列作为碼本身，或作为碼最初的 L_M-1 位的数字来使用是不可以的。

由上述条件，考虑 $d=2$ 的二元碼的結构。由 (iii)，概率最小的两个消息所对应的长度是相等的，由 (iv)，长度为 L_M 的两个代碼的最后数字为 0 及 1，其他的数字都相同。这两个代碼若对应

于第 M 号及 $M-1$ 号消息，則可把两个消息看做一个复合消息，其概率取两个概率的和。从而把这个复合消息及其他 $M-2$ 个消息合在一起改为 $M-1$ 个消息，再依照其概率大小順序排列之后，就可决定最小概率的两个消息的最后数字。这种操作可以重复施行。于是 Huffman 編碼法可以归納成下列五点：

（i） M 个消息依概率大小排列之。

（ii） 对于概率最小的两个消息，求其概率之和。

（iii） 把上面求出的两个消息的概率之和看做一个消息的概率，再与其他消息依概率大小排列之。

（iv） 继續(ii)与(iii)的操作，直到最后剩下一个以 1 为概率的消息。

（v） 对每两个联在一起的消息分别賦予二进位数，相反执行上述操作，原来 M 个消息就各获得了二元碼。对于上例的信息源，利用这种編碼法可以得到表 14.3，它确实比 Shannon 和 Fano 的方法有較高效率。

表 14.3 Huffman 編碼法

消息号数 s	消息概率 p_s		二元碼	碼 长
1	0.20		11	2
2	0.19		10	2
3	0.18		011	3
4	0.17		010	3
5	0.15		001	3
6	0.10		0001	4
7	0.01		0000	4
			平均碼长 L_{qv}	2.72

$$
\text{傳輸速率}=\frac{2.61}{2.72}=0.960(\text{必特/秒})
$$

第5章　有噪音的离散信道

有噪音的离散信道是信息論中最有兴趣的部分，从 Shannon 的研究以来得到了許多結果，并且严格化了。本章說明有噪音信道的傳輸速率，曖昧度和散布度等概念，然后导出关于通信容量的重要定理，并且闡明其計算方法和信息傳輸誤差的上限等性质。

§15 曖昧度，散布度及傳輸速率

現在考虑傳輸信号在信道中受到噪音干扰的情况。这就是說，接收机收到的信号与发射机发出的信号不一定是一致。但是所謂不一致，如果发出同一个信号，必然收到同一信号，那末这不是噪音的干扰，而是信道中的畸变。接收机对接收到的信号有一定的校正作用时，至少在原理上可以除掉畸变的影响。今后我們不考虑畸变的情形。从信息論的角度看来，最重要的是发出同一信号之后，最后不一定得到同一信号，即所謂随机对应的情况。这就是以后所述連續信道中的噪音，从物理上說，連續随机变数的电噪音，重迭在离散值的信号波形上，結果使波形打乱。理論上，可以用現在所述的和消息相同的随机过程来表示，特别是看做与消息相同的遍历过程。本章为了簡单起見，不考虑它与以前所經过的状态有关的情形。在离散信道中不是注重波形，而是发送的代碼因噪音的干扰，随机地成为另一代碼。受到噪音干扰的信道与§11所述变换器的区别是：在后者为必然的現象在这里却是随机現象了，当§11公式(11.1)，(11.2)的信道状态为β_t时，令发射机信号为x_t，接收机收到的信号为y_t，此时信道迁移到状态β_{t+1}的概率为$p_{x_t,\beta_t}(y_t, \beta_{t+1})$。特别是，本章所論的是信道上連續发

出的代碼与以前的代碼完全独立且受到噪音干扰的情况，这样的信道没有内部的記忆，这种信道只有一个状态，上述概率中参数 β_t 可以不管，信道的性质可以用一組迁移概率 $p_{x_t}(y_t)$ 表示出来。如以后各节所述，在发射机发出的信号与以前发出的信号独立的情况下，可以导出較詳細的性质，其他的情况就非常复杂了，本章主要只叙述前者。

信息源发出的消息被发射机变成信号，經有噪音的信道傳送到接收机，接收机在收到信号之后，还原成消息。在考虑这样一个通信系統时，代表消息的或称为信息的信号以及有干扰作用的噪音必須看做两个随机过程。这里有几个熵标志着这个通信系統的特征。首先是信息源的熵，亦即发射机輸出(即信道的輸入)信号的熵 $H(x)$。本章假定发射机是正規的，从而两者的熵是相等的。其次，設信道的輸出即接收到的信号的熵为 $H(y)$。在信道中以 x 表发射端，以 y 表接收端。如果沒有噪音，則得 $H(x)=H(y)$，但有噪音时等式不能成立。由发射端及接收端信号的联合概率分布，按照§5，設其联合熵为 $H(x, y)$，又假定已知发射信号时，接收信号的条件熵为 $H_x(y)$，已知接收信号时，发射信号的条件熵为 $H_y(x)$。这些熵之間的关系为

$$H(x, y)=H(x)+H_x(y)=H(y)+H_y(x). \qquad (15.1)$$

这里和以前一样，熵以单位时間計算，或用单位碼計算。

今举例說明，試考虑二元碼信道中信息的傳輸。代碼序列由“0”及“1”两个数字組成，設 0 和 1 的概率分別是 $\frac{1}{2}$，且以每秒 1000 个二元数字的速率傳送信号。就是說信息源以 1000 必特/秒的速率产生信息。假設信道因受到噪音的干扰，在接收端每 100 个数字中有一个錯誤数字。在这些情况下如何使用熵来規定有噪音信道的傳輸速率呢？而且又不能与§12 无噪音的信道沒有关系。稍加考虑，有可能把減去錯誤接收数字平均数所剩下的差 990

必特/秒看做傳輸速率，但这是不对的。因为沒有考虑在接收端究竟哪一个数字是錯誤的。如果考虑一个极端的情况就很清楚，假定噪音非常大，以至接收信号与发射信号变成完全独立。此时不論发射端发出的信号是什么，接收端收到0或1的概率都是$\frac{1}{2}$，从而接收信号約有一半可能是正确的，实际上就是根本沒有发出什么信息时，此信道的傳輸速率也有500必特/秒。就是信道切断之后，接收端还是如同擲硬币一面为0，另一面为1一样，具有完全相同的傳輸速率，結果引起了矛盾。

当然这里所能考虑的是由发出的信息量减去接收信号所損失的信息量，也就是在收到信号时与实际上发出的信号中的不确定性。这不确定性如§5所述，乃是接收信号已經知道时所对应的发射信号的条件熵，采用条件熵的合理性是可以想象的，实际上也是如此。从而有下面的定义。

定义15.1　有噪音离散信道的傳輸速率R定义为

$$R=H(x)-H_y(x), \tag{15.2}$$

与此信道有关的所有信息源中R的最大值称为这个信道的**通信容量**，以C表之。

(15.2)式中熵以时間单位或以代碼单位表达都可以，如以时間单位表达，沒有杂音时$H_y(x)=0$，与定义12.1相一致。以代碼单位表达的熵当信道沒有噪音时沒有特别的意义，在現在情况下，它表示对于1个代碼来說含有多少正确的信息。

对照一下上例，收到数字0时，实际上发出0的后驗概率为0.99，发出数字1的后驗概率为0.01。收到数字1时，把上面数字0及1的概率代换一下，于是

$$H_y(x)=-[0.99\log_2 0.99+0.01\log_2 0.01]=0.081$$

（必特/数字）。

然而1秒間发生1000个数字，所以乘1000后得81必特/秒。从

而信道的傳輸速率为 1000－81＝919 必特/秒。在第二个极端例子中，接收代碼的后驗概率都是 $\frac{1}{2}$. 所以

$$H_y(x)=-\left[\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right]=1\text{(必特/数字)},$$

或为 1000 必特/秒，因此，信道的傳輸速率为 0；結果不发生矛盾。再利用(15.1)和(5.7)式，得

$$R=H(x)+H(y)-H(x,y)=H(y)-H_x(y), \quad (15.3)$$

此式与(15.2)一起，共有三种表示式。

条件熵 $H_y(x)$ 称为有噪音信道有关的**曖昧度**❶ (equivocation)。这个量标志着接收信号的平均含糊程度。与此相对，发射信号在接收端因噪音的干扰会收到錯誤的信号，$H_x(y)$ 就表示发射信号在接收端引起錯誤的范圍，称为**散布度**(dissemination)。

§16 关于通信容量的基本定理

有噪音的离散信道，不論发的是何种信号，由于受到噪音的干扰，不能傳輸确实的信息，于是定义了通信容量，这与无噪音信道确实傳輸不同信息的最大值的概念不同。但是以冗长的形式傳輸信息，可以减少收到信号的錯誤概率。例如，把信号重复地多发几次，接收端也重复地接收，从比較中可以判定出現最多的信号为实际发出的信号，这样可使錯誤的概率变得很小。但是要求誤差概率为零的話，就要无限多次重复，結果傳輸速率会想作是零。实际上，这种印象是完全錯誤的，这由下列定理可以說明。事实上，在发射端发出适当編碼的信号，接收端收到受噪音干扰的信号后进行适当的譯碼，只要傳輸速率不超过通信容量 C，就可在这信道上以任意小的錯誤概率傳輸信息。就是說在这种情况下，通信容量

❶ 或者譯为可疑度。——校者注

是"几乎正确地"傳送不同信息的最大量，这和无噪音信道的情况下的概念在本质上是相同的。此时，要以发射机和接收机內較长的延迟时間作为代价换取小的誤差。但是以比C还大的速率而且无誤地傳送信息是不可能的。此事可用图16.1来表示。横軸是送向信道信息的发生速度$H(x)$，纵軸是信道中的曖昧度$H_y(x)$．斜綫內的点为可能实現无誤傳輸的部分。

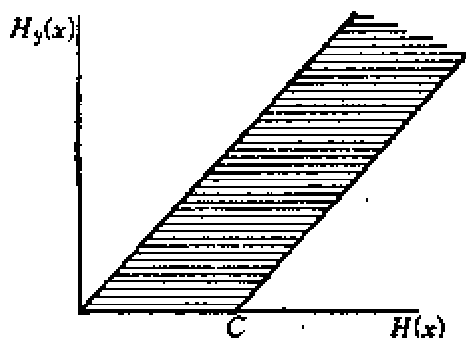


图　16.1

这个惊人的事实被Shannon首先指出[8]，后来他的定理被Feinstein[25,26]严格化了。这可以説是信息論中輝煌的成果。

定理16.1　設有噪音的无記忆离散信道❶的通信容量为C，傳輸速率为R．如果碼长m充分大，則存在一种編碼方法使当以$R<C$的傳輸速率傳送信息时，其錯誤概率可以小于$Fe^{-B(C-R)^2}$．式中F和B为信道的固有常数，与R及n无关。当$R>C$时，就不可能有象$R<C$的情况使錯誤概率达到任意小的編碼方法。

証明这个定理之前，要求有一些准备知識。

在发射端，长度为n的代碼序列一般用u表示，接收端的代碼序列以v表示。設代碼序列$u=(x_1, x_2, \cdots, x_n)$，它在发射端所取的概率为$P(u)$，則

$$\log P(u)=\sum_{i=1}^{n}\log P(x_i).$$

实际上发出u时，接收的代碼序列为v的条件概率設为$p_v(u)$，代碼序列v設为$(y_1, y_2, \cdots, y_n)$，則

$$\log p_v(u)=\sum_{i=1}^{n}\log p_{y_i}(x_i).$$

此外，u和v的集合分别以U和V表之，得

❶　所謂"无記忆离散信道"，如§15开始所述，是指信道中由噪音所生的变换与过去完全独立的信道。

$$\left.\begin{aligned} R &= H(x) - H_y(x), \\ H(x) &= -\frac{1}{n}\sum_U P(u)\log P(u), \\ H_y(x) &= -\frac{1}{n}\sum_U\sum_V p(u, v)\log p_v(u). \end{aligned}\right\} \tag{16.1}$$

这个定理在叙述三个輔助定理后再証明❶。

輔助定理 A 对于任意正数 ε 和 δ，存在一个 $n(\varepsilon, \delta)$，当任意 n 大于 $n(\varepsilon, \delta)$ 时，使公式

$$\left|H(x) + \frac{1}{n}\log P(u)\right| < \varepsilon \tag{16.2}$$

不成立的 u 的概率之和小于 δ。同样，使

$$\left|H_y(x) + \frac{1}{n}\log p_v(u)\right| < \varepsilon \tag{16.3}$$

不成立的 u 和 v 的联合分布概率之和小于 δ。

証明 由大数法則及定理 9.1 可以求得。以下只需要用不等式 $P(u) < 2^{-n(H(x)-\varepsilon)}$ 及 $p_v(u) > 2^{-n(H_y(x)+\varepsilon)}$，使这些公式不成立的概率分别以 δ^- 及 δ^+ 表示。

輔助定理 B 令 Z 表示联合概率大于 $1-\delta_1$ 的 (u, v) 的集合，設 U_0 为满足 $P(U_0) > 1-\delta_2$ 的 U 的子集。当 $u \in U$ 时，使 $(u, v) \in Z$ 的 v 的集合以 A_u 表示，設满足 $p_u(A_u) \geqslant 1-a$ 的 $u \in U_0$ 的集合为 $U_1 \subset U_0$，則 $P(U_1) > 1-\delta_2-\dfrac{\delta_1}{a}$。

証明 令 $p_u(A_u^c) > a$ 成立的 u 的集合为 U_2，此处 A_u^c 是 A_u 的余集。于是 $u \in U_2$ 时 $p(u, A_u^c) \geqslant aP(u)$，再由 A_u 的定义，$\sum_{U_1}(u, A_u^c)$ 与 Z 不相交。从而 $\delta_1 \geqslant \sum_{U_2} P(u, A_u^c) \geqslant aP(U_2)$，或 $P(U_2) \leqslant \delta_1/a$。即 $p(U_2 \cdot U_0) \leqslant \delta_1/a$，利用 $U_1 = U_0 - U_0 \cdot U_2$，得 $P(U_1) = P(U_0) - P(U_0 \cdot U_2) > 1-\delta_2-\dfrac{\delta_1}{a}$。

輔助定理 C 某一代碼序列发出之后，假定接收端判断实际上发来的代碼序列 u 具有最大后驗概率 $p_v(u)$，設錯誤的平均概率

---

❶ 以下的証明是对于二元碼的信道而言，熵的对数以 2 为底。

为 P_e，则暧昧度为

$$H_v(u) \leqslant -P_e \log P_e - (1-P_e)\log(1-P_e) + P_e \log(M-1), \tag{16.4}$$

其中 M 为发射端发出代碼序列的个数[34]。

証明　設收到代碼序列 v 时的錯誤概率为 $P_e(v)$. 于是根据定义，$p_v(u)$ 的最大値为 $1-P_e(v)$。从而

$$-\sum_U p_v(u)\log p_v(u)$$

在最大値以外的 $p_v(u)$ 都等于 $P_e(v)/(M-1)$ 时达到最大，故得

$$-\sum_U p_v(u)\log p_v(u) \leqslant -[1-P_e(v)]\log[1-P_e(v)] - P_e(v)\log\frac{P_e(v)}{M-1}, \tag{16.5}$$

此处錯誤的平均概率为

$$P_e = \sum_V P(v) P_e(v). \tag{16.6}$$

在(16.5)式的两边都乘以 $P(v)$，按 v 相加，再利用(16.6)，得

$$\begin{aligned} H_v(u) \leqslant &-\sum_V P(v)[\{1-P_e(v)\}\log\{1-P_e(v)\} + P_e(v)\log P_e(v)] \\ &+ P_e\log(M-1). \end{aligned} \tag{16.7}$$

式中右边第一項为收到代碼序列 v 时的錯誤条件熵，由 §5 之(5.8)式，条件熵小于沒有条件的熵，把第一項的条件熵換以沒有条件的熵，就得出(16.4)。

現在开始定理 16.1 的証明，并且詳細地說明这个定理的意义。此处設 $R<C$ 及已知任一正数 e，則存在一个滿足下列性质的 $n(e, R)$，当 $n \geqslant n(e, R)$ 时，长度为 n 的代碼序列至少存在 2^{nR} 个，且当发送其中一个代碼序列时，錯誤接收的平均概率必小于 e。此时接收的方法是，接收到代碼序列 v 时，可以判断实际上发出了具有最大 $p_v(u)$ 的 u。首先証明下列事实。

对于任一 $e>0$ 及 $R<C$，存在一个 $n(e, R)$，当 $n \geqslant n(e, R)$ 时，含有 n 个碼的代碼序列中，至少存在 2^{nR} 个代碼序列集合 $\{u_i\}$ 滿足下列条件：

(i) 分别对每个 u_i 說，存在滿足条件 $p_{u_i}(B_i) > 1-e$ 的 v 的集合 B_i，

(ii) B_i 彼此不重迭。又当 $R>C$ 时，不存在这样的代碼序列。

首先，在 $R>C$ 的情況下，傳輸速率超过了 C，这是矛盾的。

其次，R 达到 C 时，$H(x)$，$H_y(x)$ 分别記作 $H^0(x)$，$H_y^0(x)$。即 $C = H^0(x) - H_y^0(x)$。

如輔助定理 A 所述，对于 $p_v(u) > 2^{-n(H_y^0(x)+\theta_1)}$，$p(u) < 2^{-n(H^0(x)-\theta_2)}$，分别

規定 $n_1(\varepsilon_1, \delta_1^+)$，$n_2(\varepsilon_2, \delta_2^-)$，今后令 n 为大于 n_1，n_2 的常数。又在輔助定理 B 中的 Z 及 U_0，分别令为满足上列两个不等式的集合。此时对于任意的 $u\in U_1$（与此对应的常数 a 小于 e）以及 A_u 中的 v，得

$$\frac{p_v(u)}{P(u)} > \frac{2^{-n(H_y^u(x)+\varepsilon_1)}}{2^{-n(H^u(x)-\varepsilon_2)}} = 2^{n(C-\varepsilon_1-\varepsilon_2)}.$$

上式乘以 $P(v)$，再依照 A_u 的 v 相加，可得

$$\frac{P(u, A_u)}{P(u)} > 2^{n(C-\varepsilon_1-\varepsilon_2)}P(A_u),$$

但在左边 $p_u(A_u)\leqslant 1$，結果得出

$$P(A_u) < 2^{-n(C-\varepsilon_1-\varepsilon_2)}. \tag{16.8}$$

設 $u_1, \cdots, u_M$ 为满足下列条件的 U 的元素組成的集合 W：(a) 对于每一 u_i 存在满足 $P_{u_i}(B_i)>1-e$ 的 v 的集合 B_i。(b) $P(B_i)<2^{-n(C-\varepsilon_1-\varepsilon_2)}$ 成立。(c) B_i 互不重迭。(d) 集合 W 为最大集合，即找不到那样的 u_{M+1} 及 B_{M+1}，使集合 $u_1, \cdots, u_{M+1}$ 满足 (a) 到 (c)。

由定义，对于任何 $u\in U_1$，存在满足 $p_u(A_u)\geqslant 1-a>1-e$ 的 A_u，且 (16.8) 式成立。此外，对于任何 $u\in U_1$，$A_u-A_u\cdot\sum\limits_j B_j$ 与 B_i 沒有重迭的部分，

$$P(A_u-A_u\cdot\sum_i B_i) < 2^{-n(C-\varepsilon_1-\varepsilon_2)}.$$

若 u 不在 W 中，則

$$p_u(A_u-A_u\cdot\sum_i B_i)\leqslant 1-e, \tag{16.9}$$

即 $p_u(A_u\cdot\sum\limits_i B_i)\geqslant e-a$，或是对于一切的 $u\in U_1-W\equiv U_1-W\cdot U_1$，

$$p_u(\sum_i B_i)\geqslant e-a.$$

若取 $e\leqslant {}^1/_2$ 时，由輔助定理 B 及 W 的定义得

$$\begin{aligned} P(\sum_i B_i) &= \sum_U P(u)p_u(\sum_i B_i) \geqslant \Big\{\sum_{U_1-W\cdot U_1} + \sum_{W\cdot U_1}\Big\}P(u)p_u(\sum_i B_i) \\ &\geqslant (e-a)\Big[1-\delta_2^- - \frac{\delta_1^+}{a} - P(W\cdot U_1)\Big] + (1-e)P(W\cdot U_1) \\ &\geqslant (e-a)\Big(1-\delta_2^- - \frac{\delta_1^+}{a}\Big). \end{aligned}$$

另一方面，因 $P(\sum\limits_i B_i) < M2^{-n(C-\varepsilon_1-\varepsilon_2)}$，所以

$$M\cdot 2^{-n(C-\varepsilon_1-\varepsilon_2)} > (e-a)\Big(1-\delta_2^- - \frac{\delta_1^+}{a}\Big). \tag{16.10}$$

当 $e>\dfrac{1}{2}$ 时，也有同样的結果，以下只考虑 $e\leqslant\dfrac{1}{2}$ 的情况。今固定 $\varepsilon_1, \varepsilon_2, \delta_1^+$，

δ_2^-, a 及 n，取某一 M 值，例如 M^*，使不等式(16.10)不成立，則使(16.10)成立的 M 的最小値也比 M^* 大。例如取 $M^*=2^{nR}$, $a=e/2$, 又取 δ_1^+, δ_2^-, ε_2 适当小，n 适当大，使$\left(1-\delta_2^- - \frac{\delta_1^+}{a}\right)$大于2/3.則得 $e/3<2^{-n(C-R-\varepsilon_1-\varepsilon_2)}$，使 $C-R-\varepsilon_1-\varepsilon_2>0$ 的 ε_1, ε_2 充分小时，对于充分大的 M, $e/3<2^{-n(C-R-\varepsilon_1-\varepsilon_2)}$ 不再成立。从而必須有 $M>M^*=2^{nR}$.

其次求 e 为 n 的函数。

$$e \leqslant a+\frac{A}{B-(\delta_1^+/a)},\quad 其中\begin{cases}A=2^{-n(C-R-\varepsilon_1-\varepsilon_2)},\\ B=1-\delta_2.\end{cases}$$

为了消去 a，可作如下的考慮：上式右边当 $a=\frac{\sqrt{A\delta_1^+}+\delta_1^+}{B}$ 时达到最小值，这个值 $\frac{1}{B}[\sqrt{A}+\sqrt{\delta_1^+}]^2$ 比 $\frac{1}{B}(\sqrt{A\delta_1^+}+\delta_1^+)$ 大。此处取

$$a=\frac{\sqrt{A\delta_1^+}+\delta_1^+}{B},\quad e=\frac{1}{B}[\sqrt{A}+\sqrt{\delta_1^+}]^2$$

时，得 $a<e$. 从而

$$\frac{1}{B}[\sqrt{A}+\sqrt{\delta_1^+}]^2 \tag{16.11}$$

为对应于 e 的最小値的上限。

其次，当收到代碼序列 v 时，$p_v(u)$ 的最大値設为 $p_v(u_v)$，根据輔助定理 C 的証明，$P_e(v)$ 等于 $1-p_v(u_v)$. 此时，

$$\begin{aligned}P_e&=\sum_V P(v)P_e(v)=\sum_V P(v)\,(1-p_v(u_v))\\&=1-\sum_V P(v)p_v(u_v)\\&=1-\sum_{U_i}\sum_{B_i}P(v)p_v(u_v)-\sum_{V-\sum_{U_i}B_i}P(v)p_v(u_v),\end{aligned}$$

而 U_i 为满足 $p_{u_i}(B_i)\geqslant 1-e$ 的 u_i 的集合。設 u_0 为任意的 u_i，得

$$\begin{aligned}P_e&\leqslant 1-\sum_{U_i}\sum_{B_i}P(v)p_v(u_v)-\sum_{V-\sum_{U_i}B_i}P(v)p_v(u_0)\\&=1-\sum_{U_i-u_0}\sum_{B_i}P(v)p_v(u_v)-\sum_{V-\sum_{U_i-u_0}B_i}P(v)p_v(u_0)\\&=1-\sum_{U_i-u_0}P(u_i)p_{u_i}(B_i)-P(u_0)p_{u_0}\Big(V-\sum_{U_i-u_0}B_i\Big)\\&\leqslant 1-\sum_{U_i-u_0}P(u_i)\,(1-e)-P(u_0)\,(1-e)=e.\end{aligned} \tag{16.12}$$

由輔助定理 C 的(16.4)的右端当 $e<1/2$ 时是 e 的遞增函数，結果得

$$H_v(u)\leqslant -e\log e-(1-e)\log(1-e)+e\log(M-1), \tag{16.13}$$

此即代碼序列的曖昧度。除以 n 即得代碼的曖昧度，

$$H_y(x)=\frac{1}{n}H_v(u). \tag{16.14}$$

此处 $M=2^{nR}$，$R<C$ 或簡单地取 $e\to 0(N\to\infty)$ 时，由(16.13)得

$$H_y(x)=\frac{1}{n}H_v(u)\to 0 \quad (n\to\infty). \tag{16.15}$$

特别令 $P(u_i)=2^{-nR}$ 且当 $n\to\infty$ 时，$\frac{1}{n}[H(u)-H_v(u)]\to R$。这个事实，Shannon 曾以另外証法获得，Feinstein 曾获得更詳細的結果。在一般条件下敘述要占許多篇幅，这里只討論对称的二元碼信道。即发射端只发射 x_1 及 x_2 两种碼构成的代碼序列，接收端也只接收 y_1 及 y_2 两种碼。假設有对称性 $p_{y_2}(x_1)=p_{y_1}(x_2)$。取 ε_1 及 ε_2 充分小，設 $C-R-\varepsilon_1-\varepsilon_2>0$ 及 $H^0(x)-R-\varepsilon_2>0$ 成立。(16.11)式的 δ_1^+ 沒有一个清楚的唯一的上限。这里当 $n\to\infty$ 时，ε_1 假定比 $n^{-\frac{1}{6}}$ 更快地收斂于零，容易証明 δ_1^+ 及 e 比 $1/n$ 更快地趋于零(見文献[11]第 144～145 頁的定理)。与(16.13)对照一下，不仅(16.15)的 $H_y(x)$ 趋于零，代碼序列的曖昧度 $H_v(u)$ 本身也收斂于零。对于不对称的二元碼，一般条件下的証明，用中心极限定理还不够，要使用 Feller 定理 (W. Feller: Trans. Amer. Math. Soc., 54(1943), p.361～372)可以得到上列結果，但这里不讲了。

以上結束了基本定理的証明，但若噪音对不同代碼序列的干扰不是互相独立的，即在所謂 Markoff 过程的情况下，如以后所述对于連續信道那样的証明还未完全得到。如上所述一般可求出接收誤差平均概率的上确界，对于对称二元碼的信道，Elias 曾求出一个下限[20]。

在証明中亦表明了錯誤概率漸趋于零的編碼方法的概念。图 16.2 内，左边的点表示发射端发出的长度为 n 个碼的代碼序列，右边的点则表示接收端的代碼序列。如果发出一个代碼序列，例如 S_1，接收端在点群 R_1 中有一点为所对应的收到的代碼序列的概率几乎等于 1，R_1 的一部分点属于别的 S_i 所对应的 R_i，在这部分，判斷是否对应于 S_1 是困难的。如果适当选擇使接收端 R_1，R_2，…，R_m 相互重迭部分变小的 2^{nR} 个点 S_1，S_2，…，当 n 增大

时，点群 R_1, R_2, …彼此重迭部分的概率測度接近于零。这是定理的主要意义。設信道的通信容量为 C，假定 $R<C$，当 2^{nR} 个可以任意接近 2^{nC} 个时上列定理成立，所以錯誤概率可以任意小。这就是說，在发射端可能的 2^n 个代碼序列中选出 2^{nC} 个来使用，即在代碼序列上附加冗长度。但是为了达到这个目的，n 就必須充分大，一般地說，这要求在发射机和接收机中要有很大的延迟。更重要的是，在傳送这种代碼序列时发射机和接收机內不能发生錯誤的动作。

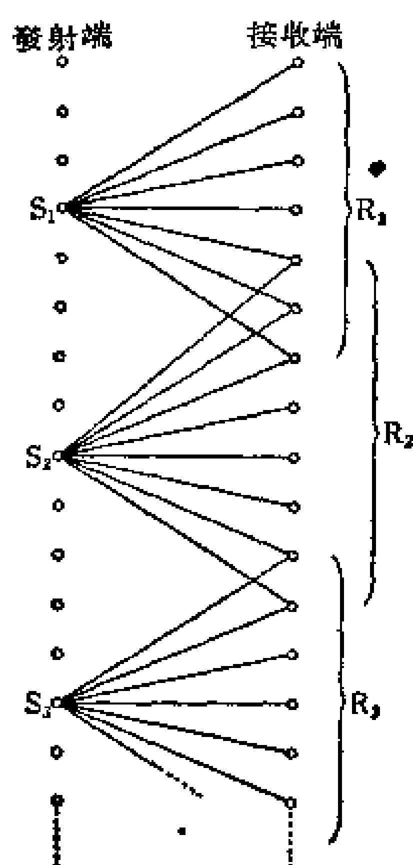


图　16.2

这个基本定理仅仅是一个存在定理，到現在为止，"具体的"最佳編碼方法还不知道。但是，各种具体的編碼方法都在研究，每一个都是当 $n\to\infty$ 时傳輸速率趋于零，这种构成方法在理論上的兴趣及在电回路上实現的实用性将在下面討論。Elias 在 1954 年最初找到对于 $n\to\infty$ 的傳輸速率为有限值的編碼，但远不能达到通信容量，只能說是半具体的东西[28]。其后，也討論了傳輸速率接近通信容量的編碼，但愈来愈缺乏具体性[27]。

这个編碼問題可以說是信息論中未解决的課題。

§17　通信容量的表現

为了討論上的方便，我們考虑信号代碼序列中的碼与过去是独立的，噪音的干扰与过去也是独立的最簡单的离散信道。受到这种噪音干扰的信道的性质，由发射端发出代碼 $i(i=1,2,\cdots,N)$

的条件下接收端收到代碼 j ($j=1, \cdots, M$) 的条件概率 $p_i(j)$ 完全規定。以 $p_i(j)$ 为 (i, j) 位置上的元素的矩陣 $(p_i(j))$ 叫做**信道矩陣**。

Shannon 虽然定义通信容量为傳輸速率的最大值，但由 $(p_i(j))$ 計算通信容量还很困难[29]，这个問題被室賀解决了[30]。信道的一个有趣性质是在发射端可能使用的代碼中选取一部分代碼傳送时，能使傳輸速率增大。这样一来，很容易想象对有噪音的信道，只要多准备通信代碼就可能增大傳輸速率，但不論 $(p_i(j))$ 如何，情形并不一定是如此，这說明了具有噪音干扰的信道的复杂性。

首先导出通信容量的計算式，从下列定理看来，傳輸速率的表示式就变得簡单了。

定理 17.1 散布度的特性方程

$$\sum_{j=1}^{M} p_i(j) X_j = \sum_{j=1}^{M} p_i(j) \log p_i(j) \quad (i=1, 2, \cdots, N) \quad (17.1)$$

的解以 $X_1, X_2, \cdots, X_M$ 表示。于是傳輸速率 R(熵/碼)为

$$-\sum_{j=1}^{M} P'_j \log P'_j + \sum_{j=1}^{M} P'_j X_j. \quad (17.2)$$

又散布度 $H_y(x)$ 可用

$$-\sum_{j=1}^{M} P'_j X_j \quad (17.3)$$

表示。其中 P'_j 为接收端出現碼 j 的概率，且有下列关系

$$\sum_{i=1}^{N} P_i p_i(j) = P'_j \quad (j=1, \cdots, M), \quad (17.4)$$

證明很簡单，由公式(17.1)及(17.4)考慮下列双綫性形式即可。恒等地有

$$\sum_{i=1}^{N} \sum_{j=1}^{M} p_i(j) X_j P_i = \sum_{j=1}^{M} P'_j X_j = \sum_{i=1}^{N} P_i \sum_{j=1}^{M} p_i(j) \log p_i(j),$$

由此即得上述定理，M 不一定与 N 相等，但上式总成立。

M 个輔助变数 $X_1, \cdots, X_M$ 及其相依的特性方程(17.1)完全規定出傳輸速率及散布度。

这儿特别取 $M=N$，且假設信道矩陣的阶数为 N。此时通信容量是在 $\sum_{j=1}^{N} P'_j = 1$ 的条件下，$P'_j \geqslant 0$，$P_i \geqslant 0$ 范圍內(17.2)式的最大值，先用 Lagrange 待定乘数法来求解极大値的問題。

$$U = -\sum_{j=1}^{N} P'_j \log P'_j + \sum_{j=1}^{N} P'_j X_j + \mu \sum_{j=1}^{N} P'_j, \tag{17.5}$$

現求 P'_j 使上式达到极大値，μ 为待定乘数。

$$\frac{\partial U}{\partial P'_j} = -(1+\log P'_j) + X_j + \mu = 0. \tag{17.6}$$

上式乘 P'_j 后，按 j 取和，則得 $C+\mu-1=0$. 其中 C 为(17.2)的极大値。因此代入(17.6)式得

$$P'_j = \exp(X_j - C), \tag{17.7}$$

但因为 $\sum P'_j = 1$，所以

$$C = \log\left\{\sum_{j=1}^{N} \exp(X_j)\right\}. \tag{17.8}$$

由(17.7)式得 P'_j 为正，但用(17.4)求出的 P_i，如以后的例所示，不一定是正的，也可能是負值。若一切值都为非負，这个 C 就是所求的通信容量，于是有下列定理。

定理 17.2　設信道矩陣 $(p_i(j))$ 为 N 阶非奇异的方陣，則有噪音信道的通信容量为 $C=\log\left\{\sum_{i=1}^{N} \exp(X_i)\right\}$，接收端碼 j 的出現概率 P'_j 为 $\exp(X_j - C)$ 时，达到最大傳輸速率。但对这些 P'_j 所求的 P_i 都是非負的。

其次我們考慮上述定理中至少有一个負的 P_i 的情形。此时在 $P_1, \cdots, P_N$ 的非負的范圍不能达到极大値，故在值域的边界上傳輸速率达到最大。从而在 $P_1, \cdots, P_N$ 中必有为零的。換言之，与零相对应的碼不能使用，此时傳輸速率变大了。今設 $P_1, \cdots, P_N$ 中只有 r 个是正的，余下来的都是零。从而信道矩陣变为

$$\begin{pmatrix} P_1(1) & \cdots & P_1(M) \\ \vdots & & \vdots \\ P_r(1) & \cdots & P_r(M) \end{pmatrix} \quad (r<M).$$

在此情况下，要求通信容量，只須求(17.2)的最大值就可以了。

于是作齐次联立方程

$$\sum_{j=1}^{M} p_i(j) S_j = 0 \quad (i=1, \cdots, r), \tag{17.9}$$

它的 $(M-r)$ 組基本解設为 $S_1^{(k)}, \cdots, S_M^{(k)}$ $(k=1, \cdots, M-r)$。求(17.2)式的最大值很困难，因而用 Lagrange 待定乘数法求其最大值，此时相当于(17.4)条件的 P'_j 必須滿足条件

$$\sum_{i=1}^{r} p_i(j) P_i = P'_j \quad (j=1, 2, \cdots, M). \tag{17.10}$$

換言之，变动 $P'_1, \cdots, P'_M$ 而求其极大值时，在变动过程中不是相互无关的。(17.10)式通常对最后解出 $P_1, \cdots, P_r$ 是必要的，其必要及充分条件是[31]

$$\sum_{j=1}^{M} P'_j S_j^{(k)} = 0 \quad (k=1, \cdots, M-r). \tag{17.11}$$

在(17.11)式的約束条件下求(17.2)的极大值，仍可用 Lagrange 待定乘数法。即求

$$U = -\sum_{j=1}^{M} P'_j \log P'_j + \sum_{j=1}^{M} P'_j X_j + \mu \sum_{j=1}^{M} P'_j + \sum_{k=1}^{M-r} \nu_k \sum_{j=1}^{M} P'_j S_j^{(k)} \tag{17.12}$$

的极大值，式中 μ, ν_k $(k=1, \cdots, M-r)$ 是待定乘数。

$$\frac{\partial U}{\partial P'_j} = -(1+\log P'_j) + X_j + \mu + \sum_{k=1}^{M-r} \nu_k S_j^{(k)} = 0, \tag{17.13}$$

乘以 P'_j 后按 j 相加可以确定 μ，并把(17.11)考虑在內得

$$P'_j = \exp\left(X_j - C + \sum_{k=1}^{M-r} \nu_k S_j^{(k)}\right), \ (j=1, \cdots, M). \tag{17.14}$$

再把此式代入(17.11)得

$$\sum_{j=1}^{M} S_j^{(k)} \exp\left(X_j + \sum_{k=1}^{M-r} \nu_k S_j^{(k)}\right) = 0, \quad (k=1, \cdots, M-r). \tag{17.15}$$

这是由 $M-r$ 个联立的超越方程中决定 $M-r$ 个待定乘数 $\nu_1, \cdots,$

ν_{M-r}。 $X_j+\sum\limits_{k=1}^{M-r}\nu_k S_j^{(k)}(j=1,\cdots,M)$为(17.1)式(但 N 可以等于 r)的一般解，由約束条件(17.15)在此一般解中决定一个。又因 $\sum\limits_{j=1}^{M}P'_j=1$，所要求的极大値为

$$C=\log\left[\sum_{j=1}^{M}\exp\left(X_j+\sum_{k=1}^{M-r}\nu_k S_j^{(k)}\right)\right].\qquad(17.16)$$

将(17.15)所决定的待定乘数代入之后，卽得所求之 C。以公式(17.10)的 P'_j 代入(17.10)，求得的 P_i 若不是負的，則所求的极大値就是通信容量，从而下列定理成立。

定理 17.3　設发射代碼有 r 个，接收代碼有 M 个，且信道矩陣的阶数为 r 的离散信道的通信容量为

$$C=\log\left[\sum_{j=1}^{M}\exp\left(X_j+\sum_{k=1}^{M-r}\nu_k S_j^{(k)}\right)\right],\qquad(17.17)$$

其中 ν_k 滿足約束条件(17.15)，又由(17.14)式的接收代碼的概率 P'_j 代入(17.10)后解出的 P_i 都是非負的。

所以由上述結果，終可求出信道的傳輸速率的最大値，卽通信容量。首先用发射端的全部 N 个代碼，由定理 17.2 求极大値，并計算出这个极大値时的 $P_i(i=1,\cdots,N)$。若其中有一个 P_i 是負的，則在发射代碼中除掉一个，以 $N-1$ 个发射代碼用定理 17.3 求极大値。将除掉的代碼一个个地改換着，求其极大値，若所对应的 $N-1$ 个 P_i 都是非負的，其极大値中最大者卽为所求的通信容量。若不論怎样选擇 $N-1$ 个代碼，对应的极大値的 P_i 总有負的，此时把发射代碼减为 $N-2$ 个。如此把发射代碼的个数降低，直到上述作法中不出現負的 P_i 为止。此时信道的最大傳輸速率，卽为所求的通信容量。若此时的发射代碼的个数为 r，則使用 r 个以下的碼，傳輸速率自然不会超 r 个碼的情況。

在以上計算过程中，可用下面的方法具体地說明(17.9)的基本解。改变变数的号数，使最初 r 阶子行列式不等于零，

$$
\left.
\begin{aligned}
&S_i^{(1)}=\frac{1}{\Delta}\begin{vmatrix} p_1(1) & \cdots & p_1(r+1) & \cdots & p_1(r) \\ \vdots & & & & \\ p_r(1) & \cdots & p_r(r+1) & \cdots & p_r(r) \end{vmatrix}, \quad (i=1,\cdots,r) \\
&\qquad\qquad\qquad\qquad \text{第} i \text{行} \\
&S_{r+1}^{(1)}=1,\ S_{r+r}^{(1)}=0,\ \cdots,\ S_M^{(1)}=0; \\
&S_i^{(2)}=\frac{-1}{\Delta}\begin{vmatrix} p_1(1) & \cdots & p_1(r+2) & \cdots & p_1(r) \\ \vdots & & & & \end{vmatrix}, \quad (i=1,2,\cdots,r) \\
&S_{r+1}^{(2)}=0,\ S_{r+2}^{(2)}=1,\ S_{r+3}^{(2)}=0,\ \cdots,\ S_M^{(2)}=0; \\
&S_i^{(M-r)}=\frac{-1}{\Delta}\begin{vmatrix} p_1(1) & \cdots & p_1(M) & \cdots & p_1(r) \\ \vdots & & & & \end{vmatrix}, \quad (i=1,2,\cdots,r) \\
&S_{r+1}^{(M-r)}=0,\ \cdots,\ S_{M-1}^{(M-r)}=0,\ S_M^{(M-r)}=1, \\
&\text{这里}\quad \Delta=\begin{vmatrix} p_1(1) & \cdots & p_1(r) \\ \vdots & & \\ p_r(1) & \cdots & p_r(r) \end{vmatrix}.
\end{aligned}
\right\} \quad (17.18)
$$

以上的討論是假定在 (17.1) 有解的情况下进行的 (若解不是唯一的，只考虑其特解就够了)。如果解不存在，則由发射端至少有 $N-\rho$ 个通信代碼不使用，其中 ρ 为信道矩陣的阶数，用上述方法便可求出通信容量。我們省略了此种情况的詳細論証，但可注意它具有一些有趣的性质[30]。

数值例 設信道矩陣为

$$
\begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}.
$$

解(17.1) 得 $X_1=4\log 2$, $X_2=X_4=-6\log 2$, $X_3=0$, 应用定理 17.2, 得 $C=\log\left(17+\frac{1}{32}\right)$, 与此对应的 P_i 为

$$\left.\begin{aligned} P_1=P_2&=\frac{16}{17+\frac{1}{32}}\left(2-\frac{1}{512}\right)>0,\\ P_3&=\frac{16}{17+\frac{1}{32}}\left(1+\frac{1}{16}-\frac{1}{512}\right)>0,\\ P_4&=\frac{16}{17+\frac{1}{32}}\left(\frac{1}{128}-4\right)<0. \end{aligned}\right\}$$

亦即 P_1, P_2, P_3 为正，P_4 为負。此处命 $P_4=0$，得信道矩陣为

$$\begin{pmatrix} 1/2 & 0 & 0 & \vdots & 1/2 \\ 1/2 & 1/2 & 0 & \vdots & 0 \\ 0 & 0 & 1 & \vdots & 0 \end{pmatrix},$$

由(17.18)式得基本解为 $S_1^{(1)}=-1$, $S_2^{(1)}=1$, $S_3^{(1)}=0$。此外，$X_1=-2\log 2$, $X_2=X_3=0$ 为一特解，約束条件(17.15)为

$$-\exp(-2\log 2-\nu)+2\exp(\nu)=0.$$

所以 $\nu=-\frac{3}{2}\log 2$，由此得出

$$e^C=2\exp\left(-2\log 2+\frac{3}{2}\log 2\right)+\exp(0)=1+\sqrt{2}.$$

所以

$$\left.\begin{aligned} P_1'&=\frac{1}{2+\sqrt{2}},\\ P_2'=P_4'&=\frac{1}{2(2+\sqrt{2})},\\ P_3'&=\frac{1}{1+\sqrt{2}} \end{aligned}\right\}$$

及

$$\left.\begin{aligned} P_1=P_2&=\frac{1}{2+\sqrt{2}},\\ P_3&=\frac{1}{1+\sqrt{2}}. \end{aligned}\right\}$$

这些 P_i 都是正的。当去掉 P_4 以外的代碼时，求出的通信容量比此种情况小。从而得到通信容量为

$$C=\log(1+\sqrt{2}) \quad (熵/碼).$$

在更一般的情况下，即当代碼序列受到信道的限制，且噪音具有 Markoff 性质时，也曾討論过通信容量的表达問題[32]。

适用于二元碼信道的本节結果，有人曾詳細計算过[33]。也有人討論了多个离散信道鏈索式联接的情况[35]。

用儿何方法表示通信容量，計算的意义更显得清楚[30]，Shannon 曾对此作过討論[58]。

第6章　有噪音的离散信道的編碼

本章將要討論在有噪音干扰下傳輸信息的情况，卽如何編碼才能使接收的誤差为最小，同时又要效率高。前章已經說明存在一种理想的編碼方法，但实际上极重要的問題是如何进行具体的計算。这里首先讲什么是編碼，多数的研究分别为两类，卽群編碼与非群編碼。

§18 編碼的概念

在信道已經給定的情况下，我們希望求出最有效傳輸信息的編碼。亦就是要使傳輸速率达到最大，同时使接收誤差概率尽量小的問題。如第5章所述，无噪音的信道沒有誤差，事情比較簡单，已得到很好的結果，但有噪音干扰时就比較复杂。对于有噪音信道的編碼，从信息論的立場来說，它的概念如下。

首先將信息源产生的消息中冗长的內容用§14的方法取消，压縮成有信息意义的消息（对于无噪音信道是一种代碼变换器），再附加了能減少接收錯誤的冗长数字之后傳送出去，就是所要的編碼法。根据§16的基本定理知道，对于有噪音的离散信道，使接收錯誤接近于零，且通信速率尽量接近于通信容量的編碼是存在的，但是如何具体构成这种編碼是一个問題。这个問題还沒有得到解决，从理論上說也是很有趣的。实际上，在发射机和接收机內都不允許有过大的延迟，把代碼按有限个加以分区再进行傳送，此种想法很实用，曾多方被构想过。

到現在为止，檢查錯誤的碼有**5中取2碼**，**7中取3碼**❶，一

❶ 使用在国际通信中。

般地説，**n 中取 m 碼及 2-5 位碼**❶ 等比較实用。自从信息論出現以来，对編碼进行了最有系統的研究。最簡单的是 0 及 1 两个数字构成的二元碼，下面以二元碼为中心进行討論。

首先来闡明能糾正或檢出錯誤的二元碼的一般原理。长度为 n 的二元碼❷全体共有 2^n 个，由此碼組中任取两个碼，設为 X_1 和 X_2，而这两碼的第 i 个数字設为 $x_{1,i}$ 和 $x_{2,i}$ $(i=1, 2, \cdots, n)$，即

$$\left.\begin{aligned} X_1 &= (x_{11}, x_{12}, \cdots, x_{1n}), \\ X_2 &= (x_{21}, x_{22}, \cdots, x_{2n}). \end{aligned}\right\} \tag{18.1}$$

此时定义

$$L(X_1, X_2)=\sum_{i=1}^{n}(x_{1i}-x_{2i})^2 \tag{18.2}$$

为 X_1 与 X_2 的**距离**(或称 **Hamming 距离**)。换言之，即是两个碼內相对应数字不一致的个数。由长度为 n 的碼的全体 2^n 个內取出 M 个，如果 M 个碼彼此間的距离比 d 大或等于 d 时，則称此 M 个碼的集合为**最短距离为 d 的碼組**(簡单地就叫碼)。在下列表內的二元碼中，保証能糾正及檢出一定数目以下的数字差錯。

d	碼的性质
1	普通二元碼
2	能够檢出任意一个位置的数字錯誤
3	能够糾正任意一个位置的数字錯誤，能够檢出两个錯誤
2γ	能够糾正 $\gamma-1$ 个錯誤，可以檢出 $2\gamma-1$ 个錯誤
$2\gamma+1$	能够糾正 γ 个錯誤，可以檢出 2γ 个錯誤

如果发生比表中所能保証糾正或檢出的个数更多的錯誤，一般説不能檢出，但也有可能檢出錯誤的产生。

这里，$d=2R+1$ 的二元碼組也容易改为 $d=2R+2$ 碼組。改

❶ 有一部分电子計算机使用着这种碼。

❷ 由 n 个"0"或"1"数字排成的碼，本章称为长度为 n 的二元碼。

法为当一个碼含有偶数个1时，追加一个"1"，其他的碼追加一个"0"，最初的碼組最短距离为$2R+1$，这样的距离只有在含有偶数个1的碼与含有奇数个1的碼相遇时才发生，同时含有偶数个1的两个碼及同时含有奇数个1的两个碼距离为偶数，变成$2R+2$以上了[34]。

但从信息傳送的立場看，这是通过发射端按一定方法在消息上附加冗长的数字，使接收端以最小的錯誤还原成原来的消息这样的操作，以减小信道中由于出現噪音干扰而使消息发生的錯誤。在此意义下，最短距离的碼組不是糾正而只是檢出錯誤个数的界限，因为不涉及到有多少个錯誤的发生概率，所以与接收誤差沒有直接关系。由一定的編碼法构成的代碼，也規定了接收端还原为原来代碼的譯碼法❶，因而若不計算接收的錯誤概率，就很难比較已知碼組的优劣。在此意义下平均距离这一概念如以后所示，比最短距离更重要❷。

在发射端，对給定的消息配列成数字，再加上冗长数字之后构成代碼，在接收端則把收到的代碼还原为原来的消息，如果有簡单的規則，那末用电子綫路来处理这些代碼是很便利的。发射端的消息已給定时，把它表成数字，再附加一些特定的数字，在接收端利用收到的代碼与有約束的M个代碼之距离作比較，其中具有距离最短的代碼判断为所发的代碼，这种**距离判断法**乃是一般的編碼及譯碼法，容易知道当消息个数很多时，所需要的电子綫路非常

---

❶ 接收以上代碼时，沒有考虑到各数字的电波形，只是判断0或1，实用上可以考虑电波形，檢出R个誤差时，把波形最不正常的数字改正之后，就可以糾正R个錯誤[36]。

若不明确指定接收机内的譯碼方法，則不能比較某一种碼組的优劣。但本章为了簡单，不考虑各个数字波形的錯誤糾正，只考虑各个数字的振幅值是大或是小，就算收到是1或是0。

❷ 这在 Slepian 論文中曾談到，但日本的研究者早已知道了。

龐大。在此意义下，下面构成的群碼是最有組織的編碼及譯碼方法。

所謂**群碼**，如(18.1)所示，卽是长度为 n 的二元碼 X_1 及 X_2 之間有下列結合法則的碼：

$$Y = X_1 \oplus X_2 = (y_1, y_2, \cdots, y_n), \tag{18.3}$$

其中 $$y_i = x_{1i} \oplus x_{2i}.$$

$$\left(\begin{array}{l}\text{此卽以 2 为模的和，換言之，}\\ x_{1i}=x_{2i}=0,\ \text{或}\ x_{1i}=x_{2i}=1\ \text{时}\ y_i=0,\\ x_{1i}=0,\ x_{2i}=1,\ \text{或}\ x_{1i}=1,\ x_{2i}=0\ \text{时}\ y_i=1.\end{array}\right)$$

如此得出的 M 个碼所成的碼組全体，构成一个 Abel 群，例如：

$$X_1 = (01011100),$$

$$X_2 = (11110000),$$

卽得 $$X_1 \oplus X_2 = (10101100).$$

这个 Abel 群的单位元为一切数字都是零的碼 $O=(0, 0, \cdots, 0)$。容易証明，此种构成 Abel 群的碼組共有 2^k 个碼（k 是整数）。

如果消息的个数恰好是 2^k 个，则用 k 位二进位数表示时，可設其第 j 个数字为 a_j。这个二进位数以下列法則附加冗长数字 a_i 而构成的碼組称为**組織碼** (systematic code)。卽 a_i 为

$$a_i = \sum_{j=1}^{k} \gamma_{ij} a_j \quad (i = k+1, \cdots, n), \tag{18.4}$$

而 γ_{ij} 为 0 或 1 的数字，矩陣 (γ_{ij}) 为組織碼的固有矩陣，这里的和 $\sum$ 如(18.3)的$\oplus$一样，是以 2 为模的加法(mod 2)。这叫做**一致监督** (parity check)。此时 $a_1, a_2, \cdots, a_k$ 叫**信息位置** (information position)，$a_{k+1}, \cdots, a_n$ 叫**监督位置**(check position)。例如(γ_{ij})为

$$\begin{pmatrix} 1011 \\ 1101 \\ 1110 \end{pmatrix}$$

时，对消息(0101)的組織碼为(0101101)，又消息(1110)的組織碼为(1110001)。

这种組織碼，在发射端当已知表示消息的信息位置的数字时，很容易添加监督位置的数字，在接收端由收到的信息位置的数字，重新以規定好的(γ_{ij})算出监督位置的数字，把它与接收碼內的监督位置的数字作比較，不論編碼譯碼都是有高度組織的，即使消息的个数很多时，也不难用电子綫路实現出来。喜安曾提出組織碼与群碼之間关系的下列定理[37, 38]。

定理 18.1 一切群碼都是組織碼，反之，一切組織碼都是群碼。

証明 滿足(18.4)式一致监督的碼組內任意两个碼，若依照(18.3)式結合之后，易知同样滿足一致监督关系，从而一切組織碼为群碼。反之，一切群碼为組織碼一事，只要說明 Abel 群可以用基底(base)表示(参看本丛书中有关群論的著作)。即由 2^k 个碼中选取 k 个碼，按(18.3)式取其一切可能的綫性組合，就可表示出其他一切的碼。从而把 2^k 个碼加以排列，可得到以 0 和 1 为元素的 2^k 行 n 列矩陣，其秩为 k。从而此矩陣有 k 个綫性独立的列，其他 $n-k$ 个列可以用此 k 列按(18.3)的法則表示出来。只取此 k 列所成的矩陣的秩为 k，有 2^k 行，k 列，且其中有 k 行是綫性独立的。但这些行皆以 2 为模 (mod 2) 形成加群，因其中 k 行綫性独立，2^k 行不含相同的元。从而此矩陣为 0 到 2^k-1 的 k 位数 2 进位展开。結果此 k 列数字是信息位置，任何位的数字可以与其他无关地选取，余下的 $n-k$ 列的数字可由 k 列数字按照(18.3)的結合取得。

組織碼中最簡単的情况为 $k=n-1$。监督位置的数字只有一位，当信息位置中 1 的个数为偶数时，监督位置的数字为 0，当出現奇数个时就为 1❶. 这个碼組的最短距离为 2. 此种情况非常簡单，几乎不会出錯誤。单是为了檢出錯誤是很适用的，此法常在电子計算机上使用。

距离判别法，或是更有組織的方法，說明如何具体地編碼及譯

---

❶ 反之，也可以偶数时为 1，奇数时为 0。

碼才能实現如 § 16 基本定理所示的通信容量的傳輸速率，这一有趣的問題至今尙未解决，甚至于还沒有超过半具体的方法（如 § 19 所示的 Hamming-Golay 編碼[28]）。

本章下面根据不同观点討論能糾正及檢出錯誤的碼組。

§ 19 无多余的組織碼組

能糾正与檢出錯誤的碼組的构成，具有整数論上有趣的性质，从信息論的观点加以引用开始出现在 Shannon 的文献中[8]，还有 Hamming 編碼，其后被 Golay 推广了[40]。

长度为 n 的二元碼全体共有 2^n 个，把其中相互保持某一指定的最短距离 d（d 为奇数）以上的 M 个碼抽出之后，所余下的 (2^n-M) 个碼中某一碼与抽出的 M 个碼中任一碼的距离都在 $\frac{d-1}{2}$ 以下，一般地說，对于任一碼都有 $\frac{d-1}{2}$ 以上距离的碼是存在的。但在特殊情况下具有上述最后性质的碼全不存在也是可能的，即 (2^n-M) 个碼都与每一个抽出去的 M 个碼之間的距离小于 $\frac{d-1}{2}$。此种碼組，Golay 称为**无多余碼**（loseless code）。

定理 19.1 长度为 $n=2^\alpha-1$（α 是正整数）的組織碼，$2^\alpha-1-\alpha$ 位的信息位置及 α 位的监督位置之間的一致监督矩陣 (γ_{ij}) [参考 18.4]，是 2^α 个 α 位数的二进数排列，若把含有一个 1 以下的二进数除掉，这个組織碼最短距离为 3，而且是无多余碼組。

在証明之前先举一例說明定理的含义，若 $\alpha=3$，三位二进位数共有 2^3 个，(000)，(001)，(010)，(100)，(110)，(101)，(011)，(111)，除前四个含有一个以下的 1 外，余下四个，依下法将二进数字纵向排列，得

$$(\gamma_{ij})=\begin{pmatrix}1101\\1011\\0111\end{pmatrix}. \tag{19.1}$$

由此矩陣,应用(18.4)得三位的监督位置和 $2^3-1-3=4$ 位的信息位置。例如对于消息(0001),(0110),分别得到組織碼(0001111),(0110110)。

証明　首先由 $2^\alpha-\alpha-1$ 位二进位数的信息位置中选取两个任意二进位数,它們中間至少有一位以上的不同数字。先設只有一位不同数字。一个二进位数为0,另一个二进位数为1。但矩陣(19.1)的纵列至少含有2个以上的1,把前一个二进位数对应的监督位置与后一个二进位数对应的监督位置的数比較一下,有两个以上的数字是不同的。信息位置的两个碼距离为1,再包括监督位置在内最短距离为3以上。其次設信息位置的两个二进位数有两位不同数字。其所对应的监督位置的数字間的距离,稍加考虑易知它等于矩陣(19.1)中两位数不同的纵列所表示的二进位数間的距离,这个距离为1以上,信息位置与监督位置合在一起保証 $d=3$。以上証明了組織碼的最短距离为3。其次証明它是无多余碼組,对于碼組中一个碼,距离为1以内的碼的个数为 $1+\binom{2^\alpha-1}{1}=2^\alpha$。但此碼組內碼的个数为 $M=2^{2^\alpha-\alpha-1}$,在此碼組內由任一碼到另一碼距离小于1的碼的个数为 $2^\alpha\cdot 2^{2^\alpha-\alpha-1}=2^{2^\alpha-1}$。另一方面,长度为 $2^\alpha-1$,且具有此长度的一切碼的个数为 $2^{2^\alpha-1}$,这与上面計算的个数是一致的,从而这个碼組是无多余碼組。亦即以长度为 $2^{\alpha-1}$ 的碼作为整个空間,将碼組內碼之間距离为1的子空間 M 个碼全部除掉,变成最紧密的碼組。

这个定理中的碼組称为 Hamming-Golay 碼組,在叙述定理时,为了簡单,使用了二元碼,若 p 为质数,则还可以推广到 p 进位数的情形[40]。Golay 曾試图推广无多余組織碼組到 $d\geqslant 5$ 的情形,其必要条件为二項式展开的最初几个系数之和 $\sum_{i=0}^{\frac{d-1}{2}}\binom{n}{i}$ 恰好是2的冪,研究結果当 $n=90$ 时,最初三項和为 2^{12},$n=23$ 时最初四項和为 2^{11},前者由群的性质不能得到无多余碼組,后者如表19.1的一致监督矩陣得出最短距离为7的組織碼。信息位置为12位,监督位置为11位。

Golay 証明 d 为5以上时,碼愈长,无多余碼組存在的可能性愈小[41],上面定理的构成法还未給出。最近 Lloyd 研究了不管是

否組織碼，碼長在 10^{10} 以下时，保持 $d=5,7,9$ 的碼除上述以外其他不存在[57]。

在定理 19.1 內 $a=3$ 的情况下，組織碼內信息位置与监督位置适当地换一下，則在接收端施行一致监督，結果能把錯誤的位数簡单地表示出来。此事由 Hamming 所发現[39]。設二元碼第 i 位数为 x_i，监督位置的数字可由信息位置 x_3, x_5, x_6, x_7 作出：

$$\left.\begin{aligned} x_1&=x_3\oplus x_5\oplus x_7,\\ x_2&=x_3\oplus x_6\oplus x_7,\\ x_4&=x_5\oplus x_6\oplus x_7.\end{aligned}\right\}\qquad(19.2)$$

表 19.1

信息位置

100111000111
101011011001
101101101010
101110110100
110011101100
110101110001
110110011010
111001010110
111010100011
111100001101
011111111111

监督位置

接收端收到二元碼 $(y_1, y_2, \cdots, y_7)$ 时，計算下列三个数字：

$$\left.\begin{aligned} \xi_1&=y_1\oplus y_3\oplus y_5\oplus y_7,\\ \xi_2&=y_2\oplus y_3\oplus y_6\oplus y_7,\\ \xi_3&=y_4\oplus y_5\oplus y_6\oplus y_7,\end{aligned}\right\}\qquad(19.3)$$

ξ_3,ξ_2,ξ_1 看做二进位数的数值时能表示出錯誤的位置。其时 (000) 表示沒有錯誤。例如发出 (1100110) 时，第三位数字因受到噪音的干扰出現为 1. 此时接收到的信号为 (1110110)，由 (19.3) 計算得 $\xi_3=0$, $\xi_2=1$, $\xi_1=1$，二进位数 011 为 3. 这表示第三个数字錯了，改为 0 就好了。

一般地說，組織碼不能使接收端的一致监督的結果立刻把錯誤的位置表示出来。如上所述，信息位置为四位，碼长为七位的值，如能适当地扩大，仍能把糾正的位置表示出来，这种特殊的組織碼很实用，应加以注意。

定理 19.1 的 Hamming-Golay 碼組，可以推广到最短距离为 3 以上的情况。但这不一定是无多余碼組，下一节的后半便来讲述这个問題。

§ 20　非无多余的組織碼組

最短距离为一定的組織碼如前节所述的无多余碼組是非常特殊的情况，以后将可看到，沒有此种性质的倒是大多数。此种不一定是无多余碼組的組織碼乃是 Muller 碼組[42, 43]及§19 Hamming-Golay 碼組的推广[44]。前者由四个研究者独立地发现[45, 46, 51]，但 Muller 討論得最詳細，首先讲述他的理論。

对于函数上 p 个互相独立的邏輯变数 $X_1, X_2, \cdots, X_p$，任意 Boole 函数 f 可以写成下列加法标准形式(参看計算机的 Boole 代数)：

$$f=\sum_{j=0}^{2^p-1} f_j X_1^{i_1} X_2^{i_2} \cdots X_p^{i_p}, \tag{20.1}$$

其中把 j 用二进位数展开(i_k 取 0 或 1)，得

$$j=\sum_{k=1}^{p} i_k 2^{k-1}, \tag{20.2}$$

且当 $i_k=1$ 时 $X_k^{i_k}=X_k$，$i_k=0$ 时 $X_k^{i_k}=\overline{X}_k$ (X_k 的补碼)。(20.1) 式的和 $\sum$ 为模 2 的加法(普通 Boole 代数都取邏輯和)。此时以最大碼 I 及 X_j 以 2 为模的和 $\overline{X}_k=I\oplus X_k$ 代替 (20.1) 中的 $\overline{X}_k$，则 f 为只用 X_k 表示的函数：

$$\begin{aligned} f=g_0 I &\oplus[g_1 X_1 \oplus g_2 X_2 \oplus g_4 X_3 \oplus \cdots \oplus g_{2^{p-1}} X_p] \\ &\oplus[g_3 X_1 X_2 \oplus g_5 X_1 X_3 \oplus g_6 X_2 X_3 \oplus \cdots \oplus g_{2^{p-2}+2^{p-1}} X_{p-1} X_p] \\ &\oplus \cdots\cdots\cdots\cdots \\ &\oplus[g_{2^p-1} X_1 X_2 \cdots X_p], \end{aligned} \tag{20.3}$$

上式为 f 的多項式展开，系数 g_i 取 0 或 1，这个展开式的演算如下。

若变数 u，v 分别表示成二进位数：

$$u=(u_0, u_1, \cdots, u_{t-1}),$$

$$v=(v_0, v_1, \cdots, v_{t-1}),$$

其中 u_i, v_i 取值 0 或 1。(20.3)中的"$\oplus$"定义为

$$u\oplus v=(u_0\oplus v_0,\ u_1\oplus v_1,\ \cdots,\ u_{t-1}\oplus v_{t-1}),$$

系数 g_i 的乘法規定为

$$g_i u=(g_i u_0,\ g_i u_1,\ \cdots,\ g_i u_{t-1}),$$

又积为 $$uv=(u_0v_0,\ u_1v_1,\ \cdots,\ u_{t-1}v_{t-1}).$$

此处 $X_1, X_2, \cdots, X_p$ 及 I 是 2^p 維矢量，X_1 为 0 及 1 的交錯排列，X_2 为 00 及 11 的交錯排列，…X_k 是 2^{k-1} 个 0 及 2^{k-1} 个 1 的交錯排列，I 是 2^p 个 1 的排列，O 则是 2^p 个 0 的排列：

$$\left.\begin{aligned} X_1&=(01010101\cdots\cdots1),\\ X_2&=(00110011\cdots\cdots1),\\ X_3&=(00001111\cdots\cdots1),\\ &\vdots\\ I&=(11111111\cdots\cdots1),\\ O&=(00000000\cdots\cdots0). \end{aligned}\right\} \qquad (20.4)$$

于是当(20.3)式中取 $g_0=g_3=1$ 以及其他的 g_i 皆为零时，$f=I\oplus X_1X_2$ 可表示为(11101110…1110)。

此时展开式(20.3)中含有 r 个以下的変数的积的諸項所成 r 次多項式中，一切以 0 及 1 为系数的此种多項式的集合称为**邏輯変数的 d 阶网** (net of order d)，而 $d=2^{p-r}$.

定理 20.1 設 $r^1, r^2, \cdots, r^t$ 为 d 阶网的元素，其中任意两个元素 r^i 及 r^j 間的距离，$i\neq j$ 时最短距离为 $d=2^{p-r}$ 以上。

証明 可用归納法来証明。$p=1$ 时，因为 $X_1=(01)$，由(20.3)得 $f=g_0I\oplus g_1X_1$，属于 2 阶网的元素 $g_0=0$, $g_0=1$ 分别对应着(00)，(11)，又一阶网的元素为(00)，(01)，(10)，(11)，因此能满足定理。其次設 $p=k$ 时，假定对一切 d 定理都成立，現在要証明对 $p=k+1$ 也成立。由(20.3)容易知道，若任意两元素 r^i 及 r^j 属于 d 阶网，则由 $r^i\oplus r^j=r^l$ 而得到的元素 r^l 也属于同一网。从而 r^i 及 r^j 間的距离以 $L(r^i, r^j)$ 表示时，得

$$L(r^i, r^j)=L(r^i\oplus r^j, O)=L(r^l, O),$$

只要証非 O 的 r^l 属于此网，且满足 $L(r^l, O)\geqslant d$ 就够了。設 r^l 为 $k+1$ 个邏

輯变数的函数，可写成

$$r^l = f_1 \oplus X_{k+1} f_2. \tag{20.5}$$

其中 f_1 及 f_2 是 k 个邏輯变数的函数。f_1 由 (20.4) 的 $X_1, X_2, \cdots, X_k$ 的性质容易知道是 $d/2$ 阶网元素(由 2^k 个数字組成) 的两个元素排列而成，f_2 为 d 阶网元素(由 2^k 个数字組成)的两个元素排列而成。現分四种情况来考虑。

(a) 設 f_2 为 O，則 $r^l = f_1$. f_1 看做 $k+1$ 个邏輯变数的函数时，可写做 $r^l = \overline{X}_{k+1} f_1 \oplus X_{k+1} f_1$. $X_{k+1} f_1$ 及 $\overline{X}_{k+1} f$ 分别为用 2^k 个数字所成的 $d/2$ 阶网的元素及 2^k 个 0 的排列，从而 $L(X_{k+1} f_1, O) \geqslant d/2$, $L(\overline{X}_{k+1} f_1, O) \geqslant d/2$, $X_{k+1} f_1$ 及 $\overline{X}_{k+1} f_1$ 在同一位置的数字不能同时是 1，故得

$$L(r^l, O) = L(X_{k+1} f_1, O) + L(\overline{X}_{k+1} f_1, O) \geqslant d. \tag{20.6}$$

(b) 若 f_1 为 O，則 $r^l = X_{k+1} f_2$,

$$L(X_{k+1} f_2, O) \geqslant d. \tag{20.7}$$

(c) 設 f_1 及 f_2 都不是 O，且 $f_1 = f_2$，則得

$$r^l = f_2 \oplus X_{k+1} f_2 = \overline{X}_{k+1} f_2, \tag{20.8}$$

于是

$$L(\overline{X}_{k+1} f_2, O) \geqslant d. \tag{20.9}$$

(d) 設 f_1 及 f_2 都不是 O，且 $f_1 \neq f_2$,

$$r^l = f_1 \oplus X_{k+1} f_2 = \overline{X}_{k+1} f_1 \oplus X_{k+1} f_3. \tag{20.10}$$

其中因为 $f_1 \neq f_2$ 所以 $f_3 = f_1 + f_2$ 不是 O，是 $d/2$ 阶网的两个元素的排列。从而 $L(\overline{X}_{k+1} f_1, O) \geqslant d/2$，又 $p = k$ 时定理成立，所以 $L(X_{k+1} f_3, O) \geqslant d/2$ 成立，于是得 $L(r^l, O) \geqslant d$.

由此定理得出具有最短距离为 2^{p-r}，长度为 2^p 的碼組。这个碼組内碼的个数，可由 d 阶网内 (20.3) 式的系数 g_i 的个数計算得

$$2^{\sum_{i=0}^{r} \binom{p}{i}}. \tag{20.11}$$

这个碼組不是前一节的无多余碼組，也不一定是最緊密的碼組。除 d 阶网的碼以外，有时存在与这些碼的距离为 d 以上的碼。$d = 2^p$, 2^{p-1}，当 p 为 4, 2, 1 时，网以外那样的碼不存在，除此以外的 d 值，当 p 增大时，多半存在那样的碼。用电子計算机檢查的結果，$p=5$, $d=8$ 时不存在，$p=6$ 时存在。对于比这定理有更多碼的編碼法，由此定理的改进，还能作出比(20.5)多一些的碼[47]。

这定理的碼是群組織碼。現在省略这个一致监督矩陣的証明。为了方便，$\sum_{i=0}^{r}\binom{p}{i}$ 个信息位置的数字以 a_0; a_1, a_2, $\cdots$, a_p; a_{12}, a_{13}, $\cdots$, $a_{(p-1)p}$; a_{123}, $\cdots$; $a_{123\cdots r}$, $\cdots$, $a_{(p-r+1)(p-r+2)\cdots p}$ 表示。亦卽在自然数 1, 2, 3, $\cdots$, p 之中选擇 r 个以下的数作为 a 的右下方足标。同样，监督位置 $b_{12\cdots(r+1)}$, $b_{12\cdots r(r+2)}$, $\cdots$; $\cdots$; $b_{123\cdots p}$ 为由自然数 1, 2, $\cdots$, p 中取 $r+1$ 个以上的数作为 b 的右下方足标，监督位置的位数共 $2^p-\sum_{i=0}^{r}\binom{p}{i}$. 其次，以 $a_\alpha^{(j)}$, $b_\alpha^{(j)}$ 表示 a 或 b 的右下足标有 j 个自然数，又当含在 α 的一切自然数都在 β 内时写作 $\alpha\subset\beta$. 而 0 則規定含在一切 β 内。今以 D_{ij}^r 表示 1 当 $\binom{j-i-1}{r-i}$ 为奇，表示 0 当 $\binom{j-i-1}{r-i}$ 为偶，在定理 20.1 的組織碼組中，(18.4) 对上述可能的 β 得表示式如下：

$$b_\beta^{(j)}=\sum_{i=0}^{r} D_{ij}^r \sum_{\alpha\subset\beta} a_\alpha^{(i)} \quad (j=r+1,\ r+2,\ \cdots,\ p). \tag{20.12}$$

此处和都是以 2 为模的加法运算，且关于 α 的和是对由构成 β 的 j 个自然数中选擇一切可能的 i 个自然数形成的 α 进行运算。例如 $\beta=1346$ 时，对于 α 的和在 $i=2$ 时得

$$\sum_{\alpha\subset\beta} a_\alpha^{(2)}=a_{13}\oplus a_{14}\oplus a_{16}\oplus a_{34}\oplus a_{36}\oplus a_{46}{}^{[48]}.$$

再例如 $p=3$ 时，設 $X_1=(01010101)$, $X_2=(00110011)$, $X_3=(00001111)$, $I=(11111111)$, 則一次多項式 $g_0I\oplus g_1X_2\oplus g_2X_2\oplus g_4X_3$ 为 $d=2^{3-1}=4$ 阶网。系数 g_0, g_1, g_2, g_4 的一切組合共 2^4 个，容易驗証，这 16 个二元碼組的最短距离为 4. 例如 $g_0=g_1=1$, $g_2=g_4=0$ 时，$I\oplus X_2$ 网的元素为 (11001100)。其次叙述一下一致监督法則。信息位置的系数 g_i 有 4 个，位数为 4；用上述的表示法若有 a_0, a_1, a_2, a_3, 則监督位置 a_{12}, a_{23}, a_{13}, a_{123} 为四位数。例如因 $D_{02}^1=D_{12}^1=1$, 所以 $b_{12}=a_0\oplus a_1\oplus a_2$. 如此得

出相当于(18.4)的一致监督矩陣为

$$\begin{pmatrix} 0111 \\ 1011 \\ 1101 \\ 1110 \end{pmatrix}.$$

发射端按照这个矩陣将监督位置的数字附加之后发出去，接收端則按同一矩陣由信息位置計算监督位置的数字且与收到的碼內的监督位置的数字相比較，以发現有沒有錯誤，这在下一节內还要詳細叙述。但是与此不同的譯碼法，所謂**多数判断法**(majority test)，可以考虑定理20.1的組織碼[49]。

Hamming-Golay 碼曾被喜安和 Slepian 推广过[44]。它不一定是无多余碼組，它是在同样长度及同样信息位置的位数中錯誤概率最小的碼組。它可以糾正两个数字的錯誤。

定理 20.2　当监督位置的位数为α，选擇信息位置的位数k为

$$k \geqslant 2^{\alpha} - \sum_{i=0}^{3}\binom{\alpha}{i}. \tag{20.13}$$

其中碼长为$n=k+\alpha$。作一致监督矩陣(18.4)如下：第1列由α个1排列而成，第2列以下为恰好含有$\alpha-1$个1的α位的二进位数共α种排列，再以下的列为含有$\alpha-2$个1的α位的二进位数共$\binom{\alpha}{2}$种排列，再以下接下去出現恰好k列α行。由这种一致监督矩陣选出的二元組織碼当$n>2^{\alpha}-1$时，出現一个数字的錯誤时有$2^{\alpha}-1$种糾正方法，若

$$2^{\alpha}-\sum_{i=0}^{3}\binom{\alpha}{i} \leqslant n \leqslant 2^{\alpha}-1, \tag{20.14}$$

则对一个数字的錯誤都能糾正，当发生两个数字的錯誤时，則有$2^{\alpha}-1-n$种糾正方法。

証明可以參看文献[38]的2.7节。在(20.14)式內，当$n=2^{\alpha}-1$时即变成上一节的 Hamming-Golay 碼。

§21 組織碼的譯碼法及小长度組織碼組的例

到前节为止，討論了特殊情况下的組織碼构成法，本节將討論这种組織碼組在接收之后能糾正或檢出錯誤的譯碼方法，并且叙述不容易納入系統理論的小 n, k 值的組織碼构成法[33]。

令长度为 n，信息位置的位数为 k 的組織碼組所属的二元碼为

$$A_1=O=(0, 0, \cdots, 0), A_2, A_3, \cdots, A_\mu,$$

其中 $\mu=2^k$. 长度为 n 的一切二元碼共 2^n 个(此群以 B_n 表示) 由(18.3)的結合法則作成一个 Abel 群，易知以上的 $A_1, A_2, \cdots, A_\mu$ 为这个群的一个子群。从而群 B_n 可以由这些子群作**陪集** (Coset, 亦称副群)的展开，即

$$B_n=\begin{Bmatrix} O, & A_2, & A_3, & \cdots, & A_\mu \\ S_2, & S_2A_2, & S_2A_3, & \cdots, & S_2A_\mu \\ S_3, & S_3A_2, & S_3A_3, & \cdots, & S_3A_\mu \\ S_\nu, & S_\nu A_2, & S_\nu A_3, & \cdots, & S_\nu A_\mu \end{Bmatrix} \qquad (21.1)$$

其中 $\mu=2^k$, $\nu=2^{n-k}$. 在以上排列中，B_n 的任意元素不会出现两次，而且 B_n 的一切元素都出現。第一行以外的行称为陪集。这个 B_n 的元素排列中第一列的元素 $S_2, S_3, \cdots, S_\nu$ 称为**陪集首項** (Coset leader)。陪集首項中的任意元素換以陪集中任一元素，仍得到相同的陪集。例如第 i 行的陪集首項为 S_i，对应的第 i 行陪集为 $S_i, S_iA_2, \cdots, S_iA_\mu$，今以 S_iA_k 代替 S_i 作为陪集首項，則产生 S_iA_k, $(S_iA_k)A_2$, $(S_iA_k)A_3$, $\cdots$, $(S_iA_k)A_\mu$，这与第 i 行陪集的元素，除順序不同之外，是相同的。从而可以取該陪集中(即这一行元素中)含 1 的个数最少的碼作为陪集首項。例如 (0000), (1100), (0011), (1111) 是最短距离为 2 的組織碼組，展开为陪集得

0000	1100	0011	1111
1010	0110	1001	0101
1110	0010	1101	0001
1000	0100	1011	0111

在陪集首項中換成含1最少的碼，得

0000	1100	0011	1111	
1010	0110	1001	0101	
0010	1110	0001	1101	(21.2)
1000	0100	1011	0111	

信道上，傳送第一行碼組中任一碼时，若因噪音的干扰，发生錯誤的个数在最短距离以下，則这个碼可以接收陪集中的碼。錯誤数字的个数比此还多时，可以轉变到碼組中的其他碼。組織碼按照(18.4)那样的一致监督法則由信息位置的数字决定监督位置的数字，作成与消息对应的碼 F 发出去。接收端收到信息位置的碼时，由 k 位数字用相同的一致监督法則与监督位置所对应的数字比較，一致时写作0，不一致时写做1。如此作成的 $n-k$ 位二进位数叫做**对应于 F 的一致监督数** (parity check number)，以 $Q(F)$ 表之。例如以下要叙述的 (21.3) 的陪集展开碼組的一致监督法則为 $a_3=a_1$, $a_4=a_1\oplus a_2$，当傳送碼(1110)时，若第二个数字錯誤地收为0，接收碼为 (1010)。由信息位置的数字計算得第三与第四个数字分别是0, 1，因为对于接收碼为1, 0，故与此碼对应的一致监督 $Q(F)$ 为 11。

定理 21.1 在作出长度为 n 的組織碼組 $O, A_2, \cdots, A_\mu$ 的陪集展开(21.2)时，若任意两个碼 F_1, F_2 属于同一行，則相同的一致监督法則作成的一致监督数 $Q(F_1)$ 及 $Q(F_2)$ 是相等的。又陪集首項 $Q(S_i)$ 可以按整数 $i-1$ 的二进位数表示那样的順序排列。例如

$$
\begin{array}{cccc}
0000 & 1011 & 0101 & 1110 \\
0100 & 1111 & 0001 & 1010 \\
0010 & 1001 & 0111 & 1100 \\
1000 & 0011 & 1101 & 0110
\end{array}
\qquad (21.3)
$$

上列陪集展开碼組的一致监督法则是 $a_3=a_1$, $a_4=a_1\oplus a_2$, 这个排列中,第二行内的碼皆适合$a_3=a_1$的一致监督,但不满足$a_4=a_1\oplus a_2$的一致监督。第二行的一致监督数为 01;第三、第四行分别为 10, 11. 但第一行即碼組内碼所对应的一致监督数自然是 00. 从而一致监督数及陪集首項之間有下列的对应关系:

$$
\begin{aligned}
00 &\to 0000=S_1\\
01 &\to 0100=S_2\\
10 &\to 0010=S_3\\
11 &\to 1000=S_4.
\end{aligned}
$$

关于这个定理的証明,首先看群 B_n 内任意两个碼結合生成的碼所对应的一致监督数,乃是由原来两个碼所对应的一致监督数以 2 为模相加而得。从而同一行的碼,亦即同一陪集的碼有相同的一致监督数。因为碼 $0, A_2, \cdots, A_\mu$ 的一致监督数为 $00\cdots0$, 陪集 S_i, S_iA_2, $\cdots$, S_iA_μ 的碼的一致监督数都是 $Q(S_i)$. 其次陪集展开 (21.2) 为有 2^{n-k} 行的群 B_n, 其信息位置有 k 位都是零的碼共 2^{n-k} 个。余下的 $n-k$ 位监督位置的数字恰好是一致监督数,所以这些碼各行分摊一个使成为陪集首項,各行作适当調整即得定理 21.1.

由此定理,就可糾正接收碼的錯誤,再按下列方法还原成原来的发射碼。假定接收碼为群 B_n 中的 F. 由一致监督矩陣的信息位置及监督位置計算出一致监督数 $Q(F)$, 于是得到陪集首項設为 S_i. 由此再作 $S_i\oplus F$, 即还原成原来的碼。換言之,当 F 出現在陪集展开(21.1)的第 i 列时,断定发射碼为 A_i. 这种譯碼法,因为是一致监督矩陣构成的碼,可以証明是錯誤最少最准确的譯碼方法。

設数字 0 正确地接收为 0 的概率为 q, 因噪音的干扰而錯誤

地接收为 1 的概率为 $p=1-q$. 对于数字 1 假定完全是对称的，即正确接收概率为 q，錯誤接收概率为 p，当陪集首項 S_i 中含有 w_i 个 1 时，这个譯碼法正确还原出原来碼的概率为

$$p_e=q^n+\sum_{i=2}^{\nu}p^{w_i}q^{n-w_i}. \tag{21.4}$$

当有 α_i 个陪集首項含有 i 个 1 时，上式可以写成

$$p_e=\sum_{i=0}^{n}\alpha_i p^i q^{n-i}. \tag{21.5}$$

此外，很容易得出

$$\nu=\sum_{i=0}^{n}\alpha_i=2^{n-k} \tag{21.6}$$

及

$$\alpha_i\leqslant\binom{n}{i}. \tag{21.7}$$

这个 α_i 含有重要的意义。若接收碼 F 与发射碼 A_i 有 S 个不同的数字，則 $A_i\oplus F=S_i$ 中含有 S 个 1. 从而 α_i 表示在这个碼內能糾正 i 重錯誤的个数。

表 21.1 列出了当 n 和 k 的值較小时，使 p_e 为最大可能值的 α_i. 为了参考，亦列出二項式的系数 $\binom{n}{i}$. 例如 $n=11$, $k=4$ 的組織碼，正确接收概率为 $p_e=q^{11}+11q^{10}p+55q^9p^2+61q^8p^3$，此即可以糾正一个錯誤及二个錯誤，三个錯誤时有 61 种方法。其次，表 21.2 示明了对于小值 n 和 k 的組織碼的一致监督方法。例如 $n=11$, $k=4$ 的組織碼，如表中所示，第 5, 6, 7, 8, 9, 10, 11 个数字为监督位置，信息位置由第 1, 2, 3, 4 个数字四位数构成。第 5 个数 $a_5=a_1\oplus a_3$，第 6 个数字为 $a_6=a_2\oplus a_4$，…. 与此表中所給的 n 和 k 值相对应的为最佳組織碼。換言之，在此表中出現一致监督法則以外，沒有其他正确接收概率 p_e 更大的組織碼。在这个表中，对于所給的 n 和 k 值，可以用电子計算机进行計算。制作这

表 21.1 α_i

n	i	$\binom{n}{i}$	$k=2$ α_i	$k=3$ α_i	$k=4$ α_i	$k=5$ α_i	$k=6$ α_i	$k=7$ α_i	$k=8$ α_i	$k=9$ α_i	$k=10$ α_i
4	0	1	1								
	1	4	3								
5	0	1	1	1							
	1	5	5	3							
	2	10	2								
6	0	1	1	1	1						
	1	6	6	6	3						
	2	15	9	1							
7	0	1	1	1	1	1					
	1	7	7	7	7	3					
	2	21	18	8							
	3	25	6								
8	0	1	1	1	1	1	1				
	1	8	8	8	8	7	3				
	2	28	28	20	7						
	3	56	27	3							
9	0	1	1	1	1	1	1	1			
	1	9	9	9	9	9	7	3			
	2	36	36	33	22	6					
	3	84	64	21							
	4	126	18								
10	0	1	1	1	1	1	1	1	1		
	1	10	10	10	10	10	10	7	3		
	2	45	45	45	39	21	5				
	3	120	110	64	14						
	4	210	90	8							
11	0	1	1	1	1		1	1	1	1	
	1	11	11	11	11		11	11	7	3	
	2	55	55	55	55		20	4			
	3	165	165	126	61						
	4	330	226	68							
	5	462	54								
12	0	1	1	1				1	1	1	1
	1	12	12	12				12	12	7	3
	2	66	66	66				19	3		
	3	220	220	200							
	4	495	425	233							
	5	792	300								

表 21.2　一致监督法则

n \ k	2	3	4	5	6	7	8	9	10
4	32 412								
5	312 42 51	412 513							
6	32 412 51 61	412 513 623	5123 6124						
7	31 41 51 612 72	413 512 6123 7123	5134 6124 7123	61 71					
8	31 41 52 62 712 812	41 512 613 723 8123	5134 6124 7123 81234	6134 7124 8123	71 81				
9	31 41 51 62 72 812 912	41 52 612 713 823 9123	5134 6124 7123 8123 9123	61345 71245 81235 91234	7134 8124 9123	81 91			
10	31 41 51 62 72 812 912 1012	41 52 63 712 813 923 10123	534 6123 7124 8134 9234 101234	61345 71245 81235 91234 1012345	71345 81245 912356 1012346	8134 9124 10123	91 101		
11	31 41 51 62 72 82 912 1012 1112	43 53 62 713 813 912 10123 11123	513 624 714 823 9134 10234 111234		713456 812456 912356 1012346 1112345	81345 912457 1012356 11123467	9134 10124 11123	101 111	
12	31 41 51 61 72 82 92 102 1112 1212	41 52 63 712 812 913 1023 11123 12123				813456 912456 20123567 11123467 11123457	91235678 1012346 1112467 1213458	10123 11124 12184	111 121

表时，Slepian 看出一个重要事实，即以 p_e 为最大的最佳組織碼，不一定是最短距离为最大的組織碼。例如，表 21.2 中 $n=7$, $k=3$ 的組織碼的最短距离为 3，但 $a_4=a_1\oplus a_3$, $a_5=a_1\oplus a_2$, $a_6=a_2\oplus a_3$, $a_7=a_1\oplus a_2\oplus a_3$ 作为一致监督法则所組成組織碼的最短距离为 4. 但是前者正确接收的概率大。从而最短距离的考虑与最佳碼的构成沒有直接关系。

§22 非群碼組

設碼长为 n，一部分冗长数字的位数为 $n-k$，它的 $n-k$ 位不一定是由一致监督法則所給与的碼，即不由(18.3)的結合法則所成的群碼是不是一种优秀的碼組还是問題。这問題还未解决，Slepian 研究了直到 $n=30$, $k=30$ 为止的 n 和 k 的組合，在这个情况下确定了最佳組織碼。

但是对于任意的 n 和 k 的最佳碼組并不保証必須是組織碼。怎样构成此种非群碼，还不很清楚，目前这种研究很不够。本节討論当最短距离被指定时的关于碼組内碼的个数的上限的 Muller 定理，并叙述 Plotkin 用整数論观点的編碼法作为結束。

首先，把 Muller 定理論証如下。

定理 22.1[50] 长度为 n，具有最短距离 d 的二元碼組含有的二元碼的个数，当 $d\geqslant n/2$ 时，不大于 $2n$. 当 $d>n/2$ 时，二元碼的个数不大于 $n+1$.

証明 任意二元碼 a 以

$$a=(a_0,\ a_1,\ \cdots,\ a_{n-1}) \tag{22.1}$$

表示时，施以坐标变换

$$A_i=(1-2a_i)/\sqrt{n}, \tag{22.2}$$

得矢量 A:

$$A=(A_0,\ A_1,\ \cdots,\ A_{n-1}). \tag{22.3}$$

即所施之变换为当 $a_i=0$ 时，$A_i=1/\sqrt{n}$, $a_i=1$ 时，$A_i=-1/\sqrt{n}$. 从而得

$$\sum_{i=0}^{n-1}(A_i)^2=1. \tag{22.4}$$

由(22.1)所表示的二元碼集合，亦即碼組以記号 U 表之，滿足(22.4)的矢量集合以 V 表之。集合 U 經(22.2)变換之后必含在 V 內。

設 U 中任意两个碼 a 和 b，經变換 (22.2) 后得 A 和 B。若 $a_i=b_i$，則 $\frac{1}{2}-\frac{n}{2}A_iB_i$ 为 0，若 $a_i\neq b_i$，則 $\frac{1}{2}-\frac{n}{2}A_iB_i$ 为 1，从而 a 和 b 之間的距离(18.2)为

$$L(a,b)=\frac{n}{2}\left(1-\sum_{i=0}^{n-1}A_iB_i\right), \tag{22.5}$$

于是关于 d 的条件 $L(a,b)\geqslant\frac{n}{2}$ 可写做 $\sum_{i=0}^{n-1}A_iB_i\leqslant 0$，而条件 $L(a,b)>\frac{n}{2}$ 可改写为 $\sum_{i=0}^{n-1}A_iB_i<0$。

此处假定能在碼組 U 內找到 γ 个碼 $r^0, r^1, \cdots, r^{\gamma-1}$，当 $j\neq k$ 时滿足条件 $L(r^j, r^k)\geqslant n/2$。再經变換(22.2)得到碼 $R^0, R^1, \cdots, R^{\gamma-1}$ 且滿足

$$\sum_{i=0}^{n-1}R_i^j R_j^k\leqslant 0. \tag{22.6}$$

这些矢量乘以酉矩陣 M，即矢量集合 V 作旋轉，此时 $\sum_{i=0}^{n-1}A_iB_i$ 保持不变，条件(22.6)仍然成立。換言之，令 $\sum_{i=0}^{n-1}A_iB_i=AB^*$ 时，因 $A'=AM$，$B'=BM$，$A'(B')^*=AM(BM)^*=AMM^*B^*$。但因 M 是酉矩陣，MM^* 为单位矩陣，故得 $A'(B')^*=AB^*$。

其次，对 V 旋轉得到 V'，使 $(R^0)'=(1,0,0,\cdots,0)$ 成立是可能的。換言之，即旋轉 M 使 R^0 变为 $(1,0,0,\cdots,0)$。首先以酉矩陣 M_1 作旋轉：

$$M_1=\begin{pmatrix} \frac{R_0^0}{\Delta} & \frac{-R_1^0}{\Delta} & 0 & 0 & 0 & \cdots & 0 \\ \frac{R_1^0}{\Delta} & \frac{R_0^0}{\Delta} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

其中
$$\Delta=\sqrt{(R_0^0)^2+(R_1^0)^2},$$

矢量 R^0 乘以 M_1 时，除 R_0^0，R_1^0 外，其他坐标不变，R_0^0 变为 Δ，R_1^0 变为 0。同

样作 M_2 使 R_0^0, R_2^0 变化, 可以使 R_2^0 变为 0. 如此得到的矩陣 $M=M_1M_2M_3\cdots M_{n-1}$ 把 R^0 旋轉成为 $(1, 0, 0, \cdots, 0)$. 今把 $(R^0)'$ 由 V' 除掉, 使 $(R^0)'$ 含在集合 W 中。又当矢量 $(-1, 0, 0, \cdots, 0)$ 含在 V' 內时,把它由 V' 除掉,使含于 W 內。于是在 V' 內剩余的矢量第一坐标分量为 $\sum_{i=0}^{n-1}(R_i^0)'=(R_0^j)'=(R_0^j)'\leqslant 0$, 是非正的。$(R^j)'$, $(R^k)'$ 在 V' 中时,則条件

$$\sum_{i=0}^{n-1}(R_i^j)'(R_i^k)'\leqslant 0$$

可写作

$$(R_0^j)'(R_0^k)'+\sum_{i=1}^{n-1}(R_i^j)'(R_i^k)'\leqslant 0, \tag{22.7}$$

$(R_0^j)'$及$(R_0^k)'$都不是正的,(22.7)的第一項非負,从而

$$\sum_{i=0}^{n-1}(R_i^j)'(R_i^k)'\leqslant 0. \tag{22.8}$$

亦即 V' 減少一維数,因为比 V 少 1 个或两矢量、在 V' 內条件(22.6)可改写为 (22.8)。当維数为 $n-1$ 时, 条件(22.4)不按原来的样子成立, V' 的不論那一个矢量乘以一常数使(22.4)成立时并不影响 (22.8) 的成立。从而最后 $n-1$ 坐标为 0 的向量由 V' 中去掉移到 W 中去, 亦即总可能使矢量规格化。对 $n-1$ 維的 V' 重复进行以上步驟,得到 V''. 在每一过程中最多使 W 增加两个矢量,但使原来的集合减少一个維数。n 次操作之后,有 γ 个矢量移到W去,所以 γ 不大于 $2n$.

定理后一半的証明与上一半的方法相同,其中使用了条件

$$\sum_{i=0}^{n-1}R_i^jR_i^k<0. \tag{22.9}$$

此时,当矢量$(1, 0, 0, \cdots, 0)$及$(-1, 0, 0, \cdots, 0)$同时移到 W 时,(22.9)不能成立,在 V 中不剩下任何矢量。从而在最后一次操作之前,每回只移到 W 中一个矢量,最后一次有两个矢量移到 W 去,V 不再剩下任何矢量。即γ不大于 $n+1$.

这个 Muller 定理不管是否由結合法則 (18.3) 构成群碼或是非群碼, 都給出了上限。二元碼为 $2n$ 时, $d=n/2$ 是整数; 从而至少可以知道碼长 n 必須是偶数。定理 20.1 的网 $r=1$ 即最短距离 $d=2^{p-1}$ 时选的碼組, 以及以下将叙述的 Plotkin 碼組, 都是本定理的实現。

其次来叙述整数論的 Plotkin 碼組的构成法[45]。其他与此类似的方法也已經知道[51]，此处以 Plotkin 方法作为例証。它与到现在为止所述的相反，在(18.3)結合法則意义下，不一定构成群碼組。

定理 22.2 当碼长为 n，最短距离指定为 $n/2$ 时，若 $n=4m$，且 $4m-1$ 为质数，则能构成含有 $8m$ 个碼的，最短距离保持为 $2m$ 的碼組。

为了篇幅的限制，此处只叙述它的构成法。

由 Fermat 定理，設 θ 为质数，且 θ 不能被正整数 a 除尽，则下列公式成立：

$$a^{\theta-1}\equiv 1 \pmod{\theta}. \tag{22.10}$$

特别当 $\theta-1$ 是满足上式的最小正指数时，a 称为 θ 的原始根。例如 $\theta=13$，2 是原始根。以 13 为模，2^s 的同余整数表如下所示：

s	1	2	3	4	5	6	7	8	9	10	11	12
2^s	2	4	8	3	6	12	11	9	5	10	7	1

此处使用整数論的定理：质数 $\theta\,(>2)$ 的原始根为 β 时，β, β^2, β^3, $\cdots$, $\beta^{\theta-1}$ 分别是与 1, 2, 3, $\cdots$, $\theta-1$ 中某一数以 θ 为模的同余数。此点由上表可以清楚地看出。

此处設 β 为 $4m-1$ 的原始根。整数 $\xi\not\equiv 0 \pmod{\theta}$，此外若存在整数 η 使 $\eta^2\equiv\xi \pmod{\theta}$，则称 ξ 为质数 θ 的平方剩余。由整数論知道，整数 $1, 2, \cdots, \theta-1$ 恰好一半是平方剩余，即 $\beta^2, \beta^4, \cdots, \beta^{4m-2}$ 为 $4m-1$ 的平方剩余。而 β, β^3, $\cdots$, β^{4m-3} 不是 $4m-1$ 的平方剩余。其次决定 z_i：当 i 是 $4m-1$ 的平方剩余，即 $\beta^2, \beta^4, \beta^6$, $\cdots$, β^{4m-2} 的某一同余数时，使 $z_i=1$，若 i 不是 $4m-1$ 的平方剩余，即 β, β^3, $\cdots$, β^{4m-3} 的某一同余数时，令 $z_i=0$。但 $i=4m-1$ 时令 $z_i=1$。其次作二元碼 a_1, a_2, $\cdots$, a_{4m-1} 之前，首先令 a_1 的第 i

个数字为 z_i 的二元碼 $(i=1, 2, \cdots, 4m-1)$. 对于 $j=2, 3, \cdots, 4m-1$, a_j 是取 a_1 的数字作 $j-1$ 位的循环而得的二元碼。最后 A_i 是由 $4m-1$ 位的二元碼 a_i 在第 $4m$ 位数字附加一个 0 而得。$4m$ 个 0 及 $4m$ 个 1 排列而成的二元碼分别以 O, I 表示, $8m$ 个二元碼 $O, I, A_1, A_2, \cdots, A_{4m-1}, A_1\oplus I, A_2\oplus I, \cdots, A_{4m-1}\oplus I$ 是保持最短距离为 $2m$ 的碼組。距离为 $2m$ 以上的証明略去了。

例如 $m=3$, $4m-1=11$ 的情况, 2 是 11 的原始根, $2, 2^2, 2^3, \cdots, 2^{10}$ 以 11 为模, 分别与 2, 4, 8, 5, 10, 9, 7, 3, 6, 1 是同余数。其 4, 5, 9, 3, 1 是平方剩余, 其他是非平方剩余。当 $i=1, 3, 4, 5, 9$ 时 $z_i=1$, 又 $i=2, 6, 7, 8, 10$ 时 $z_i=0$. 又 $z_{11}=1$. 从而得

$$a_1=(101110\quad 00101)$$
$$a_2=(011100\quad 01011)$$
$$a_3=(111000\quad 10110)$$
$$\cdots\cdots\cdots\cdots$$
$$a_{10}=(011011\quad 10001)$$
$$a_{11}=(110111\quad 00010)$$

由此得 $4m=12$ 位的最短距离为 $2m=6$ 的碼, 共 $8m=24$ 个如下:

$$O=(000000\quad 000000)$$
$$A_1=(101110\quad 001010)$$
$$A_2=(011100\quad 010110)$$
$$\cdots\cdots\cdots\cdots$$
$$A_{10}=(011011\quad 100010)$$
$$A_{11}=(110111\quad 000100)$$
$$I=(111111\quad 111111)$$
$$A_1\oplus I=(010001\quad 110101)$$
$$A_2\oplus I=(100011\quad 101001)$$
$$\cdots\cdots\cdots\cdots$$

$$A_{10} \oplus I = (100100 \quad 011101)$$

$$A_{11} \oplus I = (001000 \quad 111011)$$

这个例子中，$A_j(j=1, 2, \cdots, 11)$ 按 (18.3) 的結合法則，显然不是群碼。用其他的 m 值試驗一下，也不一定构成群，但 $m=2$ 时也可以有与定理 20.1 的碼組一致的情况。

第 7 章　連續的通信系統

到上章为止，討論了离散的通信系統，本章将概述連續通信系統的重要概念。粗略地讲，此时是把离散的变数作細微的区間分割取其极限而加以推广。严格的理論要求有抽象的高深数学，所以要求掌握基本思想方法就够了。本章沒有討論象 Wiener 預測理論那样重要的問題，关于这个問題，請参考本丛书《随机过程的应用》。

§ 23　連續信息及噪音的表示

一般地說，連續通信系統所处理的連續信息的頻帶寬度是有限制的。即表示連續信息的时間函数 $f(t)$ 的 Fourier 頻譜限制在两个值 W_1 及 W_2 之間，这样，在电通信技术上很方便；例如对于人的声音只傳送 3000 赫以下的頻譜就可能通話，在长距离通信时，一对綫路只傳送一組会話是不經济的，可以把許多会話的頻譜中心一一錯开，使第一个会話占 0 到 3000 赫，第二个会話占 f_1 到 f_1+3000 赫，…等互相不重迭地并排在一对綫路上傳送。如果我們着眼于有限制的頻譜，那末連續信息的时間函数有个有趣的性质。

定理 23.1　設时間函数 $f(t)$ 的頻譜限制在 0 到 W 赫之間，则 $f(t)$ 可以表示为

$$f(t)=\sum_{k=-\infty}^{\infty} X_k \frac{\sin \pi(2Wt-k)}{\pi(2Wt-k)}, \tag{23.1}$$

其中

$$X_k=f(k/2W),$$

$f(t)$ 只由 $\frac{1}{2W}$ 秒❶整数倍时間上所取的振幅值所完全确定。

証明 首先，$f(t)$ 的复数頻譜 $F(\omega)$ 为

$$F(\omega)=\int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt. \tag{23.2}$$

由假設知道，当 $|\omega|>2\pi W$ 时 $F(\omega)=0$，$F(\omega)$ 在 $-2\pi W$ 到 $2\pi W$ 区間上，可以展开成有下列系数 a_k 的 Fourier 級数：

$$a_k=\frac{1}{4\pi W}\int_{-2\pi W}^{2\pi W} F(\omega)\exp[-i(\omega k)/2W]d\omega. \tag{23.3}$$

由于 $F(\omega)$ 是 $f(t)$ 的 Fourier 变換，所以进行反变换后，得 $f(t)$ 为

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t}d\omega$$
$$=\frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} F(\omega)e^{i\omega t}d\omega.$$

若令 $t=k/2W$，得

$$f\left(\frac{k}{2W}\right)=\frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} F(\omega)\exp(i\omega k/2W)d\omega. \tag{23.4}$$

比較(23.3)及(23.4)式，可得

$$a_k=\frac{1}{2W}f\left(\frac{-k}{2W}\right). \tag{23.5}$$

从而，$\cdots-(2/2W)$，$-(1/2W)$，0，$1/2W$，$2/2W$，$\cdots$ 各样本点上 $f(t)$ 的值确定时，就决定了系数 a_n，由此系数又决定了頻譜 $F(\omega)$，$F(\omega)$ 反过来对于每一 t 的值又决定了 $f(t)$。所以不含 W 以上的頻率，且已知与 $1/2W$ 秒有关点上的振幅值时，可以唯一地确定这个函数。其次，已知振幅时，反过来求函数如下：

$$F(\omega)=\sum_k a_k\exp(i\omega k/2W),\quad |\omega|<2\pi W \text{ 时},$$
$$=0 \qquad\qquad |\omega|>2\pi W \text{ 时}.$$

取其反变換，得

$$f(t)=2W\sum_k a_k\frac{\sin\pi(2Wt+k)}{\pi(2Wt+k)}$$
$$=\sum_k f\left(\frac{k}{2W}\right)\frac{\sin\pi(2Wt-k)}{\pi(2Wt-k)}. \tag{23.6}$$

---

❶ $\frac{1}{2W}$ 秒間隔有时称为 Nyquist 間隔。

即 $f(t)$ 是以样本点为中心的，峰值与样本点上 $f(t)$ 的值相等的 $\sin x/x$ 型函数列之和。

这个定理在 1935 年 Whittaker 在"内插函数論"(Interpolatary Function Theory) 书内已經有了表示法，最初論述信息波形与頻譜关系的为染谷[53] 和 Shannon[8], [53], [54]，因此亦叫做染谷-Shannon 的**样本定理**❶。由上列定理知道，頻带由 0 到 W 赫的时間函数，只考虑它在样本点上的值就够了。换言之，样本点上的值 $X_k=f\left(\frac{k}{2W}\right)$ 可以考虑为无限維空間的 $f(t)$ 的坐标。已知这些坐标时，$f(t)$ 就可看成在此空間內的一个已知点。特別是，如果时間 T 以外的 X_k 都是零，那末函数 $f(t)$ 实质上被限定在时間 T 以內。在此情况下，除 $2TW$ 个坐标以外其他的都是零，因此，頻带为 W 及时間限制在 T 以內的函数，相当于 $2TW$ 維空間內的一个点。

信息論不是考虑某一特定的函数，而是考虑它們的集合。問題在于集合中每一函数以什么概率出現，此时其坐标 X_k 考虑为随机变数。給定了 T 及 W 的函数集合，$n(=2TW)$ 維空間的概率分布密度 $p(x_1, x_2, \cdots, x_n)$ 就确定了。若函数不限定在时間 T 內，即无限长的情况下，則可考虑 $\cdots x_{-1}, x_0, x_1, \cdots$ 无限个样本点，$f(t)$ 如 § 7 所述可看做一个随机过程。但在現在的情况下，x_t 不是离散值而是連續值。此时遍历过程可看成离散情况的推广，这个問題在这里不深入討論了[12]。本章只考虑作为遍历过程的时間函数的信息。

至現在为止，时間函数 $f(t)$ 被考虑为表示信息的消息或信号，

---

❶ 苏联科学院院士 В. А. Котельников 曾在 1933 年首先提出这个定理，現在我們常把它叫做 Котельников 样本定理 [В. А. Котельников, О пропускной способности 'эфира' и проволоки в электросвязь, Матерналы к 1-му Всесоюзному с'езду по вопросам реконструкции дела связи. ВЗК. 1933 г.]。——校者注

同样也可以考虑为噪音。从 0 到 W 赫带宽的对信道有干扰作用的噪音，实际上就是在 0 到 W 赫对信道有干扰影响的宽带噪音。从而噪音 $n(t)$ 完全可按定理 23.1 展开，只考虑 $1/2W$ 秒有关的样本点的序列就够了。此时噪音可看做随机过程，其概率分布經过时間 t 后不变时，卽随机变数 x_{h+t}, x_{h+t+1}, $\cdots$, $x_{h+t+n-1}$ 的多維概率分布密度与 h 无关时，此过程叫**平稳过程**。平稳过程对任一 n 說都是 Gauss 分布时，这种噪音叫 **Gauss 噪音**。特別是，随机变数 x_t, x_{t+1}, $\cdots$ 彼此独立，且都是同一 Gauss 分布时，则叫做**白噪音**(white noise)。这些都是电通信中最常見的噪音，白噪音可以通过綫性滤波器变成 Gauss 噪音。

§ 24 連續信道的傳輸速率及通信容量

現在对于带宽为 W 赫，时間为 T 秒的信息时間函数的集合，来定义它的熵。假設 n 个相連样本点上的振幅值 x_1, x_2, $\cdots$, x_n 的 n 維概率密度为

$$P(x_1,\ x_2,\ \cdots,\ x_n). \tag{24.1}$$

連續随机变数的熵如 § 6 所述，以維数 n 除之，

$$H' = -\lim_{n\to\infty}\frac{1}{n}\int\cdots\int P(x_1,\ \cdots,\ x_n)\log P(x_1,\ \cdots,\ x_n)\,dx_1\cdots dx_n$$

$$(\text{熵/自由度}) \tag{24.2}$$

叫做**1 自由度的熵**。今用 T 代替 n 去除，則得到每秒的熵，用 H 表示。因为 $n=2TW$，所以 $H=2WH'$。

在白噪音的情况下，只考虑一維概率密度就够了，这个概率密度是中心为 0，方差为 N 的 Gauss 分布，由 § 6 的(6.17)可以知道

$$H' = \log\sqrt{2\pi eN} \tag{24.3}$$

及

$$H = W\log 2\pi eN. \tag{24.4}$$

反之，已知均方功率为 N 时，白噪音具有最大的熵。

今考虑連續随机变数的随机过程的熵。在离散变数的情况下，熵与充分大的代碼序列的对数有密切关系，同样，現在的熵与样本点充分多时的概率密度的对数有关。即当概率密度 $P(x_1, \cdots, x_n)$ 对于一切 $x_1, x_2, \cdots, x_n$ 連續时，取任意小的 δ 及 ε，除了比 δ 小的概率集合，不論怎样选 $x_1, x_2, \cdots, x_n$，对于充分大的 n，总有

$$\left| \frac{\log P(x_1, \cdots, x_n)}{n} - H' \right| < \varepsilon. \tag{24.5}$$

以下我們来考虑連續的信道。信息源发出的信号是連續时間函数 $f(t)$ 的集合，这个信号受到噪音的干扰而后收到。此处发射信号和接收信号的頻带都限定为 W 赫，时間限为 T，这种信号被 $2TW$ 个样本点所規定，其統計性质由 $2TW$ 維概率分布所規定。从而发射信号的統計性质，被其 n 維概率密度

$$P(x_1, \cdots, x_n) = P(x) \tag{24.6}$$

所决定（为了簡單起見，以單个 x 表示一組 $x_1, \cdots, x_n$）。对于噪音的性质，可以换一观点来看，当发射端发出 n 个相連样本点上的振幅序列 $x_1, x_2, \cdots, x_n$ 时，这个連續信道的統計性质就决定了接收端收到振幅值 $y_1, y_2, \cdots, y_n$ 的条件概率密度：

$$p_{x_1, \cdots, x_n}(y_1, \cdots, y_n) = p_x(y). \tag{24.7}$$

从而連續信道的傳輸速率与离散信道的情况一样，可以定义为

$$R = H(x) - H_y(x), \tag{24.8}$$

其中 $H(x)$ **为发射信号的熵**，$H_y(x)$ **为曖昧度**。于是**通信容量** C，定义为一切可能的信息源联接于信道时 R 的最大值。亦即当信号 $f(t)$ 取有限維近似时，变化 $P(x)$ 使下式达到最大：

$$-\int P(x) \log P(x) dx + \iint P(x) p_y(x) \log p_y(x) dx\, dy. \tag{24.9}$$

設 $P(x, y)$ 为 x 和 y 的联合概率密度，則

$$p_y(x)=\frac{P(x, y)}{P'(y)}, \tag{24.10}$$

其中
$$P'(y)=\int P(x, y)dx.$$

再利用

$$\iint P(x, y)\log P(x)dx\,dy=\int P(x)\log P(x)dx, \tag{24.11}$$

于是通信容量可写为

$$C=\lim_{T\to\infty}\max_{P(x)}\frac{1}{T}\iint P(x, y)\log\frac{P(x, y)}{P(x)P'(y)}dx\,dy. \tag{24.12}$$

如§6所述，熵 $H(x)$ 和 $H_y(x)$ 是与坐标系有关的值，但其差 R 的值与此无关，是一定的。

在計算 $H(x)$ 和 $H_y(x)$ 时，对数取任何底都可以。若取2为底，則与离散的情况一样，C 可以考虑为最小曖昧度的信道中每秒傳送的二进位数的最大位数。还可以如下考虑。將信号的空間分割成充分多的微小区間，使条件概率密度 $p_x(y)$ 与 x 和 y 的大小无关，在各微小区間上实质上相当于一常量。从而在連續信道的情况下，可以认为在微小区間上取离散的值，这就和离散信道一样了，于是这种情况的一些想法，本质上都适用于現在的情况。

由于接收信号的振幅被发射信号及噪音的振幅唯一地决定，所以它們的概率密度之間，存在相当于离散信道的(17.4)公式，即

$$\int P(x)p_x(y)dx=P'(y). \tag{24.13}$$

§25 具有相加性噪音的連續信道

現在叙述实际上非常重要的連續信道，在这个信道中干扰通信的噪音与信号是相互独立的，且接收信号的振幅 y 可以表示为发射信号振幅 x 及噪音振幅 n 之和，即所謂**相加的**情况。为了簡单起見，假定不論是信号或噪音，各个样本点的振幅值互相独立，

且各样本点具有相同的振幅分布，这样，代替多維概率密度，可以只考虑一維的概率密度。在相加且独立的情况下，条件概率密度 $p_x(y)$ 为噪音的大小，即接收信号与发射信号的瞬时振幅之差 $n=y-x$ 的函数，

$$p_x(y)=k(y-x)=k(n), \tag{25.1}$$

由此計算散布度 $H_x(y)$，完全是 n 的函数，以 $H(n)$ 表示。

定理 25.1 設信号与噪音互相独立，且接收信号的振幅为发射信号及噪音的振幅之和，則这个連續信道的傳輸速率为

$$R=H(y)-H(n), \tag{25.2}$$

即接收信号的熵与噪音的熵之差。又其通信容量为

$$C=\max_{P(x)} H(y)-H(n). \tag{25.3}$$

証明很簡单，因为

$$\begin{aligned} H_x(y) &= -\iint P(x)p_x(y)\log p_x(y)\,dx\,dy \\ &= -\iint P(x)k(n)\log k(n)\,dn\,dx \\ &= H(n). \end{aligned}$$

从而由(6.8)的等式，得

$$R=H(x)-H_y(x)=H(y)-H(n).$$

$H(n)$ 与 $P(x)$ 是独立的，要 R 为最大，只要使 $H(y)$ 达到极大值。但 $P(x)$ 与 $P'(y)$ 之間有(24.13)的关系，由此式变化 $P(x)$ 或 $P'(y)$，使 $H(y)$ 达到最大，此时 $P(x)$ 及 $P'(y)$ 必是非負的。当发射信号集合受到某些条件限制时，也可把这个限制考虑在内，使 R 达到最大值。

以下假定信道受到白噪音的干扰，且考虑平均发射功率保持为常数 P 的实际通信的典型情况。发射信号的平均功率为

$$\int x^2P(x)\,dx=P, \tag{25.4}$$

噪音的平均功率为

$$\int n^2k(n)\,dn=N. \tag{25.5}$$

但白噪音的概率分布密度为

$$k(n)=\frac{1}{\sqrt{2\pi N}}\exp(-n^2/2N). \tag{26.6}$$

于是計算接收信号的平均功率，利用(24.13)及$y=x+n$，得

$$\begin{aligned}\int y^2P'(y)dy&=\int y^2\int P(x)p_x(y)dx\,dy\\&=\iint(x+n)^2P(x)k(n)dx\,dn\\&=\int x^2P(x)dx+\int n^2k(n)dn\\&\quad+2\int xP(x)dx\int nk(n)dn,\end{aligned} \tag{25.7}$$

由(25.6)可知(25.7)最后一項为零，所以接收信号的均方功率为$P+N$. 但由§6的(V)，均方値固定时，給出最大熵的概率密度为正态分布，所以具有平均功率$P+N$的接收信号有白噪音的概率分布时，最大熵(每秒)为

$$H(y)=W\log 2\pi e(P+N). \tag{25.8}$$

其中信道的頻帯寬度为W赫。又由(6.17)式乘$2W$后，可得噪音的熵为每秒

$$H(n)=W\log 2\pi eN. \tag{25.9}$$

当接收信号有上述的白噪音时，給出发射信号的振幅分布是必要的，此事并不困难，由(24.13)得$P'(y)$及$p_x(y)$都是正态分布，可导出$P(x)$亦为正态分布。即发射信号也有白噪音就够了。从而利用定理25.1得出下列結果。

定理25.2　設发射信号的均方功率为P，若在频带寬度为W赫的連續信道中，有均方功率为N的白噪音干扰，则此信道的通信容量为

$$C = W \log \frac{P+N}{N} \text{（bit/秒）}. \qquad (25.10)$$

由此定理，和离散信道的情况一样，經过适当編碼之后，可以任意小的錯誤概率，以每秒 $W \log_2 \frac{P+N}{N}$ 必特的速率傳送信息。比这速率还大，要无誤地傳送信息是不可能的。为了实現这种傳輸速率，要求发射信号必須有白噪音同样的概率密度[54][55]。下面討論一个接近理想速率的通信系統。由白噪音取出 M 种不同的时間序列，每一序列的时間长度为 T. 由发射端发出这些序列时，因为在接收端 M 种样本时間序列是已知的，与实际上收到的序列逐一比較，时間 T 内两序列的振幅差的均方値最小者可以判断为实际发出的时間序列。最初选的不同的时間序列的个数为 M，M 与接收时容許的接收錯誤的概率値 ε 有关，只要由白噪音选擇序列，几乎一切情况下都有

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \frac{\log M(\varepsilon, T)}{T} = W \log \frac{P+N}{N}, \qquad (25.11)$$

虽然 ε 任意选擇，但只要 T 充分大，总可能傳送 M 种不同时間序列。上式中对数的底为 2 时，时間 T 内可以傳送 $TW \times \log_2 \frac{P+N}{N}$ 必特。Rice 对这种想法作了詳細計算，图 25.1 为其一例。在接收錯誤概率 p 为 0.5, 0.01, 0.00001 的情况下，傳輸速率 R 在充分大的 T 时都逐漸接近于 C。由白噪音可选 M 种时間序列，因为其中必含有接收誤差概率大者，所以不能讲是最佳的情况。又以接收时間序列为基准时，寻找振幅差的均方値最小的操作，与离散情况的距离判定法（§ 18）同样很麻煩。就是說当 T 增大时，比較的时間序列的个数是依照指数函数增大，这是很不实际的。在这意义下，与 § 16 的关于离散信道的通信容量的定理一样，这个定理也是一个存在定理。

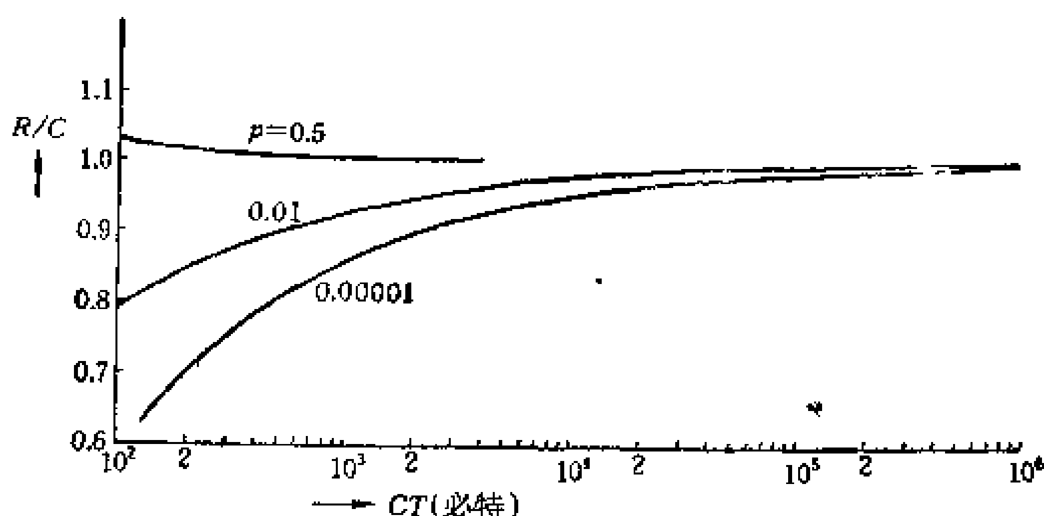


图 25.1　$N/P=0.1$ 情况下，錯誤接收概率 p 固定时，信号的时間长度 T 愈大，傳輸速率 R 愈接近于通信容量 C 的曲線。

对这个定理，也找到了与如上的概率观点不同的解释[9]。但是沒有做到象 Feinstein 那样严密的程度。

定理 25.2 是討論白噪音干扰的情况，对于任意型噪音概率密度的連續信道的通信容量定理，室賀曾細致地討論过[56]。限于篇幅，这里只录其結果如下：

定理 25.3　設信号及噪音是互相独立的，且接收信号的振幅可以用发射信号及噪音的振幅表示，其均方功率为 P，则频带寬度为 W 赫的連續信道的通信容量可由下式表示：

$$C=2W\left[-H(n)+\frac{1}{2}\log 2\pi e(P+N)\right]\text{熵/秒}, \qquad (25.12)$$

其中 $H(n)$ 为干扰这个信道的噪音熵，n_0 为噪音的平均值 $\int nk(n)dn$，N 是以 n_0 为中心的噪音的均方功率 $\int(n-n_0)^2k(n)dn$。又决定这个通信容量的接收信号的概率密度是正态分布，用

$$P'(y)=\frac{1}{\sqrt{2\pi(P+N)}}\exp\left(-\frac{(y-n_0)^2}{2(P+N)}\right) \qquad (25.13)$$

表示。但这个 $P'(y)$ 对应的发射信号的概率密度 $P(x)$ 由

$$\int P(y-n)\,k(n)\,dn = P'(y) \tag{25.14}$$

可以求得，在 x 的整个区間上是非負的。

由此定理，可以証明以前的定理 25.2 为一个特殊的情况。

参 考 文 献

[1] E. C. Cherry: A history of the theory of information, PIEE, III (1951), 383～393.

[2] 喜安善市：調制是什么？科学(电子学特集号)，25 (1955).

[3] S. O. Rice: Mathematical analysis of random noise, B. S. T. J, 23 (1944), 282～332; 24 (1945), 46～156.

[4] N. Wiener: The interpolation, extrapolation and smoothing of stationary time series (John Wiley, 1949).

[5] Y. W. Lee & J. B. Wiesner: Correlation functions and communication application, Electronics, 23 (1950), 86～92.

[6] N. Wiener: Cybernetics (John Wiley, 1947)(中譯本：維納，控制論，科学出版社)。

[7] R. A. Fisher: Contributions to mathematical statistics (John Wiley, 1950)，第 11 篇，710 頁；第 26 篇，43～44 頁；第 26 篇，47 頁.

[8] C. E. Shannon: A mathematical theory of communication, B. S. T. J. 27 (1948), 379～423, 623～658. 另有一本內容相同的单行本：C. E. Shannon & W. Weaver: The Mathematical Theory of Communication (Univ. of Illinois Press, 1949).

[9] W. G. Tuller: Theoretical limits on the rate of transmission of information, Proc. of IRE, 37 (1949), 468～478.

[10] M. J. E. Golay: Bits & Binits, Proc. of IRE (1954), 1452.

[11] W. Feller: An Introduction to Probability Theory and its Applications, I (John Wiley & Sons, 1950).

[12] J. L. Doob: Stochastic Processes (John Wiley, 1953).

[13] S. Muroga: On the capacity of a discrete channel, II, Jour. of Phys. Soc. Japan, 11 (1956), 1109～1120.

[14] Fletcher Pratt: Secret and urgent (Bobbs-Merrill, New York, 1942).

[15] K. Kunisawa: The mathematical foundation of Shannon's information source and its application to binary coding, Rep. Stat. Appl. Res., JUSE, 2, No. 1, 1～23.

[16] 本多波雄:有限时間的通信理論,通信学会杂志, 37(1954), 725～731.

[17] G. A. Barnard, III: Statistical calculation of word entropies for four Western languages, Trans. IRE, IT-1 (1955), 49～53.

[18] C. E. Shannon: Prediction and entropy of printed English, B. S. T. J. 30 (1951), 50～64.

[19] R. M. Fano: Transmission of Information, Research Laboratory of Electronics, MIT, Technical Report No. 65 (1949).

[20] 國澤清典,本多波雄,池野信一: URSI 資料(1954). 池野信一: 等长編碼,通信研究所成果报告,第 418 号 (1953).

[21] D. A. Huffman: A method for the construction of minimum-redundancy codes, Proc. IRE, 40 (1952), 1098～1101.

[22] 蒲生秀也: 电气三学会連合大会, 515(昭 28 年)喜安善市,室賀三郎: 电气通信学会杂志, 36 (1953), 498.

[23] R. M. Fano: The transmission of information, MIT, RLE Tech. Rep. No. 149 (1950).

[24] P. Elias: Predictive coding, Proc. IRE, IT-1, No. 1 (1955), 16～23.

[25] A. Feinstein: A New basic theorem of information theory, Trans. IRE. PGIT-4 (1954), 2～22.

[26] A. Feinstein: Error bounds in noisy channels without memory, Trans. IRE. IT-1 (1955), 13～14.

[27] P. Elias: Coding for noisy channels, IRE, Conv. Record, Part 4 (1955). 37～46.

[28] P. Elias: Error-free coding, Trans. IRE, PGIT No. 4 (1954), 30～37.

[29] Symposium of Information Theory (London, Ministry of Supply, 1950), 103.

[30] S. Muroga: On the capacity of a discrete channel, I. Jour. Phys. Soc. Japan, 8 (1953), 484～494.

[31] 藤原松三郎：矩陣及行列式，岩波全书（岩波书店），67頁。

[32] S. Muroga: *On the capacity of a discrete channel*, II. Jour. of Phys. Soc. Japan, 11 (1956), 1109～1120.

[33] R. A. Silverman: *On binary channels and their cascades*, Trans. IRE, IT-1. No. 3 (1955), 19～27.

[34] R. M. Fano: Lecture note of *information theory*, MIT.

[35] 三根久：有噪音重复信道的通信容量，电气通信学会全国大会，昭 31 年 11 月。

[36] R. A. Silverman & M. Balser: Coding for constant-data-rate systems—Part I, A new error-correcting code, Proc. IRE, 42 (1954), 1428～1435.

[37] Kiyasu-Zen'iti: Research and Development Data, No. 4 Elec. Comm. Lab., Nippon Tele. Corp. Tokyo (1953). 喜安善市：电气三学会东京支部連合大会，9～12 (1952, 10 月)及 22(1953, 10 月)。

[38] D. Slepian: A class of binary signaling alphabets. BSTJ, 35 No. 1 (1956), 203～234.

[39] R. W. Hamming: Error detecting and error correcting codes, BSTJ, 29 No. 2 (1950), 147～160.

[40] M. J. E. Golay: Notes on digital coding, Proc. IRE, 37, No. 6 (1949), 657.

[41] M. J. E. Golay: Binary Coding, Trans. IRE, PGIT-4 (1954), 23～28.

[42] D. E. Muller: Metric properties of Boolean Algebra and their application to switching circuits, Univ. of Illinois, Digital Computer Laboratory Internal Report, No. 46 (1953).

[43] D. E. Muller: Application of Boolean Algebra to switching circuits design and error detection, Trans. IRE, EC-3 (1954), 6～12.

[44] D. Slepian: A note on two binary signaling alphabets, Trans. IRE, IT-2, No. 2 (1956), 84～86.

[45] M. Plotkin: Binary codes with specified minimum distance, Univ. of Pennsylvania, Moore School of Elect. Eng. (1951).

[46] 三谷尙正：逐次計算器发出的数的傳輸，电气三学会东京支部連合大

会(昭26年11月)。

[47] 室賀三郎, D. E. Muller: 檢出多个錯誤編碼的构成法之改善,电气三学会东京支部連合大会(昭31年)。

[48] S. Muroga: Expression of Muller's codes in a relation matrix of Hamming's type, Univ. of Illinois 的备忘录(1954)及文献(44).

[49] I. S. Reed: A class of multiple-error-correcting codes and the decoding scheme, Trans. IRE, PGIT-4 (1954), 38～49.

[50] D. E. Muller: An upper bound for the number of certain error correcting codes, Univ. of Illinois, Digital Computer Laboratory Internal Report, No. 58 (1954).

[51] R. Paley: On orthogonal matrices, J. of Math. & Phys., XII (1933), 311～320.

[52] 染谷勳: 波形傳送(修教社,东京, 1949).

[53] B. M. Oliver, J. R. Pierce, C. E. Shannon: The philosophy of PCM., Proc. IRE, 36 No. 11 (1948), 1324～1331.

[54] C. E. Shannon: Communication in the Presence of noise, Proc. IRE, 37 No. 1 (1949), 10～21.

[55] S. O. Rice: Communication in the presence of noise—Probability of error for two encoding schemes, BSTJ, 29, No. 1 (1950), 60～93.

[56] 室賀三郎: 有噪音連續信道的通信容量, 电气通信學会杂志, 38 (昭30年), 961～969, 以及 Trans. IRE, IT-3, No. 1 (1957), 44～51.

[57] S. P. Lloyd: Binary Block Coding, BSTJ, 36, No. 2 (1957), 517～535.

[58] C. E. Shannon: Geometrische Deutung einiger Ergebnisse bei der Berechnung der Kanalkapazität, NTZ, 10, No. 1 (1957), 1～4.

校 后 記

陆志剛

信息論是最近十多年来才形成和发展的一門新学科，它在数学領域內与概率論有密切的关系，在其他科学和技术領域內，又是新兴的重要学科。

信息是事物表現的一种普遍形式，說一句話、写一封信、通一个电話、无綫电定位、导彈制导、电子計算机的运算，甚至于生命現象中的感覚过程、遺傳过程等等，都有信息，都是信息傳遞的过程[1][2][3]。信息論就是在信息是可以測度的基础上研究最有效最可靠傳遞信息的理論。由于理論的普遍性和理論中蘊藏着大量新的数学問題，亦因为最有效最可靠的要求是許多学科中普遍感兴趣的問題，所以近年来，信息論引起了許多数学家和其他科学家、工程师們的注意，信息論的概念和結果亦广泛地应用到通信[4][5][6]、电視、雷达[7][8]、导航、制导、数据处理、可靠的电子計算机运算、光学[9][10]、生物学[11]、心理学[12]、語言学等等領域中去。

网絡与信息論专家喜安善市和信息論专家室賀三郎两人合写的这本书就是介紹信息論中最基本的由 Shannon 所奠基的信息傳遞理論。虽然本书写得很簡短、扼要，然而仍能使我們对基本的問題获得一般的了解。

这本书是 1957 年以前写的，从 1957 年以来，信息論又有了許多新的进展[13]。Shannon 本人对他原来的随机編碼給出了更詳細更精确的結果[14]，其后发表了对于連續信道的研究結果[15]。給 Shannon 信息論以严格数学基础的 Feinstein，于 1958 年写成了专书[16]。应該指出，苏联数学家 Колмогоров[17][18]、Хинчин[19][20]

等人在 Shannon 理論上有着深远的工作，苏联学者将 Shannon 基本問題提得更有系統，对基本概念作出更严格的結构。捷克数学家 Nedoma[21]，Perez[22]，Winkelbauer[23] 等研究了信息傳遞更一般的模型，并以統計判决理論观点提出和处理了信息傳遞問題。最近，Добрушин[24] 介紹了苏联在信息論方面的成就和指出了許多未解决的問題，这篇文章是值得一讀的。

为了最有效和最可靠地傳遞信息，編碼和譯碼是极端重要的方法。針对有效性，有有效編碼，針对可靠性，有抗干扰編碼。Shannon 的基本定理只是一个有效編碼的存在定理，他証明傳輸速率小于通信容量时，錯誤概率随碼長 N 的增加可以变成任意小。这个証明不是結构性的，由于 N 很大，不能用于实际来构成一个实用的編碼。所以，象 Fano 編碼或 Huffman 編碼那种实用又有效的編碼方法还值得继續进行研究[25]。

自从 1950 年 Hamming 和 1956 年 Slepian 等人的工作发表以后，国外对抗干扰編碼研究得很多。实际应用时，还要考虑实現上的方便和信道中各个碼发生的錯誤不一定是独立的等情况。近年来，Green 和 San Soucie[26] 研究了一种容易实現的糾正多个錯誤的編碼方法，Calabi 和 Haefeli[27]，Abramson[28] 以及 Hagelbarger[29] 等人研究了能糾正爆发性錯誤的編碼方法。对于長度較大的碼，1954 年 Reed[30] 曾发表过 Reed-Muller 碼，1958 年 Perry[31] 实現了長度为 128 位的 Reed-Muller 碼，信息位置 64 位，监督位置 64 位，能糾正 7 个及 7 个以下的錯誤。其后 Bose 和 Chaudhuri[32] 发現了比 Reed-Muller 碼更有效的碼。在編碼方面，代数学理論用得很多，这方面可参考最近 Peterson[33] 的一本书。在实用上，序列譯碼[34] 亦值得注意。

1957 年以后，在实际应用中又提出参数随时間作緩慢变化的信道，或在傳輸过程中，参数虽然不变，但不知道它的准确值只知

道概率分布的所謂随机参数的信道，这是一个有趣的新問題，近年来也对这种信道定义了通信容量和証明了編碼定理[35][36][37]。所以修改 Shannon 的通信系統模型，使更适合于实际情况，也是极有意义的工作。

本书的第一章序論内，作者提到了 Wiener 的工作，但在正文内，一点也沒有介紹 Wiener 的过滤与預測理論。目前，有許多人把 Wiener 理論也列入为信息論，称之为 Wiener 信息論。我們认为 Wiener 理論在工程实踐上非常重要，应該予以極大的重視[38]。

近年来，有关信息論及其应用的文章非常多，有兴趣的讀者可以在[39] [40] 内找到更多的参考文献。

参考文献

[1] L. Brillouin, Science and Information Theory, Academic Press, New York, 1956.

[2] Cherry, "On human communication: a Review, a survey, and a criticism", MIT Tech. Press, and John Wiley and Sons Inc., New York, 1957.

[3] S. Kullback, Information Theory and Statistics, John Wiley and Sons Inc. Publishers, 1959 第一章.

[4] D. A. Bell, Information Theory and its Engineering Applications, Sir Isaac Pitman and Sons, London, 1st ed. (1953), 2nd ed. (1956).

[5] P. M. Woodward and I. L. Davies, "Information Theory and inverse probability in Telecommunications", Proc. IEE. vol. 99 Part III, 1952, 37～44.

[6] D. Middleton, An Introduction to Statistical Communication Theory, McGraw-Hill Co. New York 1960.

[7] P. M. Woodward, Probability and Information Theory, with Applications to Radar, Pergamon Press, Ltd., London, 1953.

[8] H. Davis, "Radar problems and information theory", IRE Conv. Rec. Part 8, 1953, 35～71.

[9] 田村稔,"情报理論と光学",电子工业,第七卷第四号,1958,1～3頁.

[10] D. Gabor, "Optical Transmission", 3rd London Symposium on Information Theory, 1955.

[11] H. P. Yockey (ed), Symposium on Information Theory in Biology, Pergamon Press, 1958.

[12] H. Quastler (ed), Information Theory in Psychology, The Free Press, 1955.

[13] P. Elias, "Information Theory and Coding", Jour. of Research of the National Bureau of Standards, Vol. 64 D No. 6, Nov-Dec. 1960, 671～673.

[14] C. E. Shannon, "Certain results in coding theory for noisy channels," Information and Control, Vol. 1.

[15] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel", Bell System Tech. Jour. Vol. 38., 1959.

[16] A. Feinstein, Foundations of Information Theory, Mc Graw-Hill Co. New York, 1958.

[17] А. Н. Колмогоров, "On the Shannon theory of information transmission in the case of continuous signals", Trans. IRE, IT-2 Dec. 1956, 102～108.

[18] А. Н. Колмогоров, Теория передачи информации, ИЗД. АН СССР 1957.

[19] А. Я. Хинчин, "Понятие Энтропия в теории вероятностей", Успехи Мат. Наук, 8, 1953, 3～20.

[20] А. Я. Хинчин, "Об основных теоремах теории информации", Успехи Мат. Наук, 11, 1956, 17～75.

[21] J. Nedoma, "The Capacity of a discrete channel", Trans. of the first Prague Conference in Information Theory, Statistical Decision Functions, Random Processes, Prague, 1957, 143～182.

[22] A. Perez, "Sur la théorie de l'information et la discernabilité dans les problèmes de décision statistique", Trans. of the second Prague Conference in Information Theory, statistical Decision Functions, Random Processes, 1959, 413～497.

[23] K. Winkelbauër, "Communication Channels with Finite Past History", Trans. of the second Prague Conference in Information Theory, statistical Decision Functions, Random Processes, 1959, 685～831.

[24] Р. Л. Добрушин, "Математические проблемы Шанноновской теории оптимального кодирования информации", Проблемы передачи информации X, 1961.

[25] R. M. Fano, Transmission of Information, John Wiley and Sons, 1961.

[26] J. H. Green Jr. and R. L. San Soucie, "An error-correcting encoder and decoder of high efficiency", Proc. IRE 46, 1958, 1741.

[27] L. Calabi and H. G. Haefeli, "A class of binary systematic codes correcting errors occurring at random and in bursts", Trans. IRE, IT-5, 1959, 79.

[28] N. M. Abramson, "A class of systematic codes for nonindependent errors", Trans. IRE, IT-5, 1959, 150.

[29] D. W. Hagelbarger, "Recurrent codes: easily mechanized burst-correcting binary codes", Bell System Tech. Jour. 38, July 1959, 969～984.

[30] I. S. Reed, "A class of multiple-error-correcting codes", Trans. IRE. IT-4, 1954, 38.

[31] K. E. Perry, "An error-correcting encoder and decoder for phone line data", IRE Wescon Conv. Rec. Part 4, 1958, 21.

[32] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error-correcting binary group codes".

[33] W. W. Petersen, Error-Correcting Codes, MIT Press and John-Wiley and Sons, Inc., New York 1961.

[34] J. M. Wozencraft and B. Reiffen, Sequential Decoding, MIT Press and John-Wiley and Sons, Inc., New York 1961.

[35] В. И. Сифоров, "On the capacity of channels with random parameter fluctuation", IRE Wescon Conv. Rec. 1957, 107～109.

[36] В. И. Сифоров, "О пропускной способности каналов связи со случайными изменениями поглощения", Радиотехника. 13, No. 5, 1958, 7～18.

[37] Blackwell, Breimann, and Thomasian, "The capacity of a class of channels", Ann. Math. Stat. 30, 1959, 1229.

[38] Y. W. Lee, Statistical Theory of Communication, John Wiley and Sons, Inc., New York, 1960.

[39] F. L. H. M. Stumpers, "A Bibliography on Information Theory, Communication Theory-Cybernetics", Trans. IRE. IT-2. Nov. 1953; first Supplement, IT-1, No. 2. Sept. 1955; Second Supplement, IT-3, No. 2. June 1957.

[40] P. E. Green, Jr., "A Bibliography of Soviet Literature on Noise, Correlation and Information Theory", Trans. IRE. IT-2. June 1955, 91～94.